Primitive Quantum Gate Realizations of Multiple-Controlled Toffoli Gates

Mozammel H A Khan
Department of Computer Science and Engineering
East West University
Aftabnagar, Dhaka 1212, Bangladesh
mhakhan@ewubd.edu

***Abstract*—Multiple-controlled Toffoli gates are extensively used in quantum and reversible circuits. However, very limited attempt has been made for primitive quantum gate realizations of these gates. In this paper, we present realization of three-qubit Toffoli gate requiring five primitive quantum gates. We then propose realization of four-qubit Toffoli gate requiring 13 primitive quantum gates and one ancilla input. Finally, we propose realizations of (n + 1)-qubit (n > 3 control lines) Toffoli gates using ($\lceil n/2 \rceil$ + 1)-qubit and ($\lfloor n/2 \rfloor$ + 1)-qubit Toffoli gates and ancilla inputs, which make a trade off between quantum cost and number of ancilla inputs. Our proposed method outperforms the best result so far presented in the literature in terms of quantum cost and number of ancilla inputs.**

***Index Terms*—Multiple-controlled Toffoli gate, quantum circuit, quantum gate, realization of Toffoli gate, reversible circuit.**

I. Introduction

Quantum computation has become a promising model of computation by improving the complexity class of several challenging problems [1]. A survey of quantum algorithms with superpolynomial speedup is given in [2]. Experimental implementation of Shor's quantum number-factoring algorithm is done in [3-7] and that of quantum walk is done in [7-9]. Physical quantum circuit realizations in the current quantum computation technologies limit the number of quantum bits (qubit) to a small number. On the other hand, experimental errors and decoherence introduce errors during computation, which requires to reduce the number of primitive quantum gates in the circuit. Therefore, in realistic design of quantum circuits, both number of qubits and number of primitive quantum gates need to be reduced. However, in quantum circuit design process, reduction of number of qubits increases the number of primitive quantum gates and reduction of number of primitive quantum gates require more qubits. Therefore, a tradeoff between these two factors is essential. Quantum circuits are first designed at the logic level as a reversible circuit using NOT, Feynman, and multiple-controlled Toffoli gates and then the logic level reversible gates are mapped into their quantum primitive gate realizations [1]. Apart from the quantum circuit realizations, Toffoli gates are used for reversible realizations of both reversible and irreversible Boolean functions [10-19]. Therefore, efficient realizations of multiple-controlled Toffoli gates are very important for realistic design of quantum and reversible circuits.

In this paper, we show realizations of three-qubit (two control lines) Toffoli gate using five primitive quantum gates and four-qubit (three control lines) Toffoli gates using 13 primitive quantum gates and one ancilla input. We then address efficient realizations of (n + 1)-qubit (n > 3 control lines) Toffoli gates using ($\lceil n/2 \rceil$ + 1)-qubit and ($\lfloor n/2 \rfloor$ + 1)-qubit Toffoli gates with a trade off between number of primitive quantum gates (quantum cost) and number of ancilla inputs. Our proposed method outperforms the best result so far presented in the literature [20] in terms of quantum cost and number of ancilla inputs.

The rest of the paper is organized as follows. In Section II, we introduce reversible and quantum gates typically used in the design of quantum and reversible circuits. We present primitive quantum gate realization of multiple-controlled Toffoli gates with a tradeoff between number of primitive quantum gates (quantum cost) and the number of zero-initialized ancilla inputs in Section III. Finally, in Section IV, we conclude the paper.

II. Reversible/Quantum Gates

The one-qubit NOT gate is shown in Fig. 1, where the input is always complemented at the output. The NOT gate is a self-inverse gate, since $\overline{(\overline{x})} = x$. Therefore, cascaded application of two NOT gates acts as a one-qubit identity gate restoring the input at the output. The NOT gate is a quantum realizable primitive gate [1].

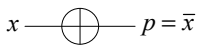


Fig. 1. NOT gate.

The two-qubit Feynman gate is shown in Fig. 2. When the control input x = 1, then the target input t is complemented at the output. Otherwise, the target input is passed unchanged at the output. The Feynman gate is a self-inverse gate, since $x \oplus (x \oplus t) = t$. Therefore, cascaded application of two Feynman gates acts as a two-qubit identity gate restoring the inputs at the

outputs. The Feynman gate is a quantum realizable primitive gate [1, 21, 22].

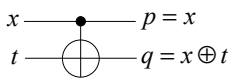


Fig. 2. Feynman gate.

The three-qubit Toffoli gate is shown in Fig. 3. The target input t is complemented at the output when the control inputs $x = 1$ and $y = 1$. Otherwise, the target input is passed unchanged at the output. The three-qubit Toffoli gate is a self-inverse gate, since $xy \oplus (xy \oplus t) = t$. Therefore, cascaded application of two three-qubit Toffoli gates acts as a three-qubit identity gate restoring the inputs at the outputs. The three-qubit Toffoli gate is not directly quantum realizable, it is realized using the two-qubit quantum realizable gates [23]. Primitive quantum gate realization of three-qubit Toffoli gate is given in [23], which requires five primitive quantum gates.

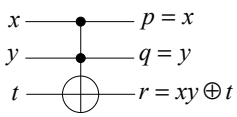


Fig. 3. Three-qubit Toffoli gate.

A generalization of the Toffoli gate with more than two inputs, called multiple-controlled Toffoli gate is frequently used in the design of quantum and reversible circuits [3-19]. A multiple-controlled Toffoli gate with n ($n > 2$) control lines is shown in Fig. 4. A multiple-controlled Toffoli gate is also a self-inverse gate. We will propose technique of primitive quantum gate realization of multiple-controlled Toffoli gates in Section III.

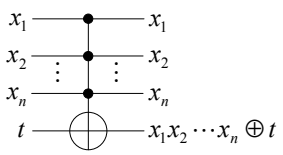


Fig. 4. Multiple-controlled Toffoli gate.

Fig. 5(a) shows the controlled-V (or CV) gate. The unitary transform $\mathbf{V} = \begin{bmatrix} (1+i)/2 & (1-i)/2 \\ (1-i)/2 & (1+i)/2 \end{bmatrix}$ is applied on the target input y at the output when the control input $x = 1$, otherwise the target input y is passed unchanged at the output. The CV gate is called the square-root-of-NOT gate, since $\mathbf{V} * \mathbf{V} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, which is the unitary transform of NOT gate. Therefore, cascaded application of two CV gates acts as a Feynman gate. Figure 5(b) shows the controlled-V^+ (or CV^+) gate. The unitary transform $\mathbf{V}^+ = \begin{bmatrix} (1-i)/2 & (1+i)/2 \\ (1+i)/2 & (1-i)/2 \end{bmatrix}$ is applied on the target input y at the output when the control input $x = 1$, otherwise the target input y is passed unchanged at the output. The CV^+ gate is also called the square-root-of-NOT gate, since $\mathbf{V}^+ * \mathbf{V}^+ = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$. Therefore, cascaded application of two CV^+ gates also acts as a Feynman gate. CV and CV^+ gates are inverse of each other, since $V * V^+ = V^+ * V = I$. Therefore, cascaded application of one CV and one CV^+ gates acts as a two-qubit identity gate restoring the inputs at the outputs. The CV and CV^+ gates are quantum realizable primitive gates [1, 21, 22].

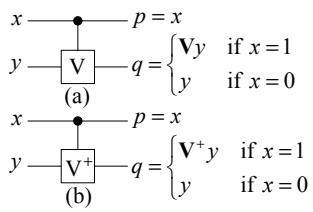


Fig. 5. Two-qubit square-root-of-NOT gates.

III. Realizations of Multiple-controlled Toffoli Gates

The three-qubit Toffoli gate can be realized as shown in Fig. 6 consisting of five primitive quantum gates, which is the same as that proposed in [23] with some rearrangement of the CV gates. The *compute part* of Fig. 6 computes the target output and the *restore part* restores the control inputs. The Feynman gate of the restore part is the mirror of the Feynman gate of the compute part.

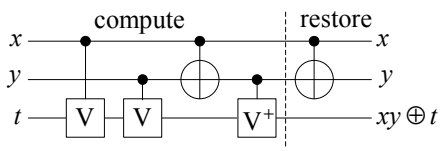


Fig. 6. Realization of three-qubit Toffoli gate.

A four-qubit Toffoli gate can be realized using three-qubit Toffoli gates and Feynman gates as shown in Fig. 7. The first Toffoli gate (from left) produces x_1x_2, which is applied as the control of the first CV gate. The first Feynman gate produces $x_1x_2 \oplus x_3$, which is applied as the control of the CV^+ gate. The control line x_3 is applied as the control of the second CV gate. The second Feynman gate is the mirror of the first Feynman gate, which restores x_1x_2. The second Toffoli gate is the mirror of the first Toffoli gate, which restores the ancilla input. When $x_1x_2 = 1$ and $x_3 = 1$, then both the CV gates are active producing NOT effect along the target line t. In this case $x_1x_2 \oplus x_3 = 1 \oplus 1 = 0$ making the CV^+ gate inactive. Thus, NOT operation is implemented along the target line t. When $x_1x_2 = 0$ and $x_3 = 1$ or $x_1x_2 = 1$ and $x_3 = 0$, then one of the CV gates is active and the other is inactive. In this case $x_1x_2 \oplus x_3 = 0 \oplus 1 = 1$ or $x_1x_2 \oplus x_3 = 1 \oplus 0 = 1$ and the CV^+ gate is active making an identity effect along the target line t. When $x_1x_2 = 0$ and $x_3 = 0$, then both the CV gates are inactive. In this case, $x_1x_2 \oplus x_3 = 0 \oplus 0 =$

0 causing the CV^+ gate inactive. Thus, the target line t will remain unchanged. Thus, the four-qubit Toffoli gate is realized. The Toffoli gate of the compute part is replaced by primitive quantum gates of the compute part of Fig. 6 and the mirror Toffoli gate of the restore part is replaced by the mirrors of the compute part (reversing the gate sequence and replacing each gate by its inverse). The primitive quantum gate realization of four-qubit Toffoli gate is shown in Fig. 8. The quantum cost of the compute part is eight and that of the restore part is five totaling quantum cost of 13. The realization requires one ancilla input.

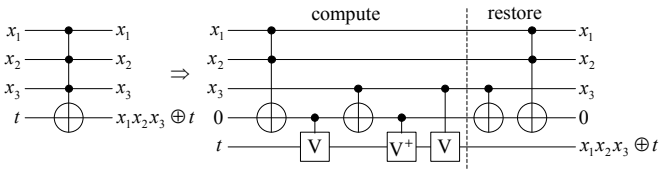


Fig. 7. Structure of realization of four-qubit Toffoli gate.

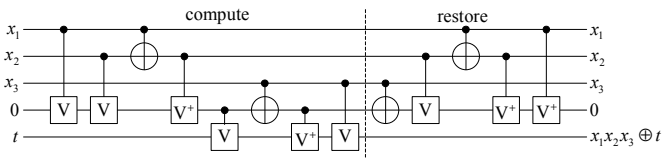


Fig. 8. Primitive quantum gate realization of four-qubit Toffoli gate.

A $(n + 1)$-qubit $(n > 3)$ Toffoli gate can be realized using two $(\lceil n/2 \rceil + 1)$-qubit Toffoli gates and two $(\lfloor n/2 \rfloor + 1)$-qubit Toffoli gates as shown in the general structure of Fig. 9. The first Toffoli gate (from left) produces $x_1 \cdots x_{\lceil n/2 \rceil}$, which is applied as the control of the first CV gate. The second Toffoli gate produces $x_1 \cdots x_{\lceil n/2 \rceil} \oplus x_{\lceil n/2 \rceil+1} \cdots x_n$, which is applied as the control of the CV^+ gate. The third Toffoli gate is the mirror of the first Toffoli gate and produces $x_{\lceil n/2 \rceil+1} \cdots x_n$, which is applied as the control of the second CV gate. The fourth Toffoli gate is the mirror of the second Toffoli gate and restores the ancilla input. When $x_1 \cdots x_{\lceil n/2 \rceil} = 1$ and $x_{\lceil n/2 \rceil+1} \cdots x_n = 1$, then both the CV gates will be active producing NOT effect along the target line t. In this case, the applied control of the CV^+ gate will be $1 \oplus 1 = 0$ making it inactive. Thus, NOT operation will be implemented along the target line t. When $x_1 \cdots x_{\lceil n/2 \rceil} = 1$ and $x_{\lceil n/2 \rceil+1} \cdots x_n = 0$ or $x_1 \cdots x_{\lceil n/2 \rceil} = 0$ and $x_{\lceil n/2 \rceil+1} \cdots x_n = 1$, then one of the CV gate will be active and the other will be inactive. In this case the applied control of the CV^+ gate will be $1 \oplus 0 = 1$ or $0 \oplus 1 = 1$ and the CV^+ gate will be active making an identity effect along the target line t. When $x_1 \cdots x_{\lceil n/2 \rceil} = 0$ and $x_{\lceil n/2 \rceil+1} \cdots x_n = 0$, then both the CV gates will be inactive. In this case, the applied control of the CV^+ gate is $0 \oplus 0 = 0$ causing it inactive. Thus, the target line t will remain unchanged. Thus, the $(n + 1)$-qubit $(n > 3)$ Toffoli gate is realized. Toffoli gates are replaced by primitive quantum gates of the compute part of their realizations and the corresponding mirror Toffoli gates are replaced by the mirrors of the compute part of their realization. The number of primitive quantum gates in the realization of the first $(\lceil n/2 \rceil + 1)$-qubit Toffoli gate will be equal to the number of primitive quantum gates of its compute part and the realization of its mirror (the third Toffoli gate) will also require the same number of primitive quantum gates. The number of primitive quantum gates in the realization of the first $(\lfloor n/2 \rfloor + 1)$-qubit Toffoli gate (the second Toffoli gate) will be equal to the number of primitive quantum gates of its compute part. In addition, two CV gates and one CV^+ gate are needed in the compute part of Fig. 9. Thus, the quantum cost of the compute part of Fig. 9 is ($2 \times$ quantum cost of the compute part of $(\lceil n/2 \rceil + 1)$-qubit Toffoli gate + quantum cost of the compute part of $(\lfloor n/2 \rfloor + 1)$-qubit Toffoli gate + 3). The second $(\lfloor n/2 \rfloor + 1)$-qubit Toffoli gate (the fourth Toffoli gate) is the mirror of the first $(\lfloor n/2 \rfloor + 1)$-qubit Toffoli (the second Toffoli gate) and requires the same number of primitive quantum gates in the realization. Thus, the quantum cost of the restore part of Fig. 9 is equal to the quantum cost of the compute part of $(\lfloor n/2 \rfloor + 1)$-qubit Toffoli gate. Therefore, the total quantum cost of Fig. 9 is ($2 \times$ quantum cost of the compute part of $(\lceil n/2 \rceil + 1)$-qubit Toffoli gate + $2 \times$ quantum cost of the compute part of $(\lfloor n/2 \rfloor + 1)$-qubit Toffoli gate + 3). Both the $(\lceil n/2 \rceil + 1)$-qubit Toffoli gate and the $(\lfloor n/2 \rfloor + 1)$-qubit Toffoli gate will require their own ancilla inputs. In addition to that one more zero-initialized ancilla input will be required in Fig. 9. Thus, the total number of ancilla inputs will be equal to (ancilla for $(\lceil n/2 \rceil + 1)$-qubit Toffoli gate + ancilla for $(\lfloor n/2 \rfloor + 1)$-qubit Toffoli gate + 1). Based on the general structure of Fig. 9, primitive quantum gate realizations of five-qubit and six-qubit Toffoli gates are shown in Fig. 10 and Fig. 11, respectively. The method of Fig. 9 is a structured and recursive method and can be used for any $(n + 1)$-qubit $(n > 3)$ Toffoli gate. The quantum costs and the numbers of ancilla qubits for up to $n = 20$ are summarized in Table I.

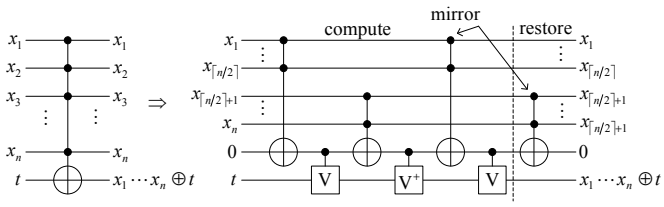


Fig. 9. General structure of realization of $(n + 1)$-qubit $(n > 3)$ Toffoli gate.

Very limited attempt has been made for primitive quantum gate realizations of multiple-controlled Toffoli gates. The best attempt for primitive quantum gate realizations of multiple-controlled Toffoli gates is presented in [20]. Comparisons of quantum cost (QC) and ancilla inputs (AI) of our work with those of [20] is shown in Table II. From Table II, we see that for $n = 3$, we achieved 7.1% QC improvement over [20] without any further improvement of AI. For $n = 4$ to $n = 12$, we achieved 3.1% to 30.0% QC improvement and 25.0% to 50.0% AI improvement over [20]. For $n = 13$ to $n = 15$, our QC is 3.7% to 14.6% more than those of [20], but we achieved 36.4%

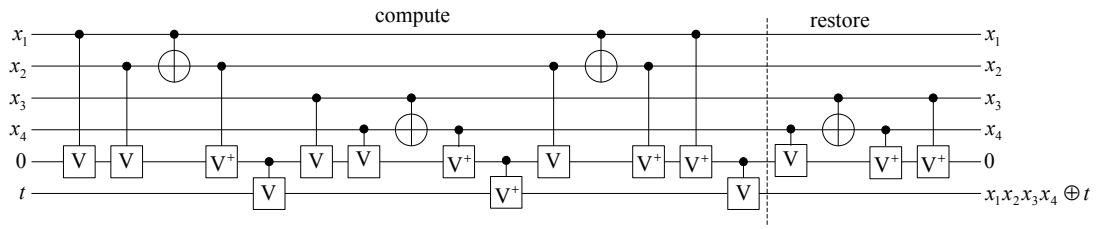


Fig. 10. Primitive quantum gate realization of five-qubit Toffoli gate.

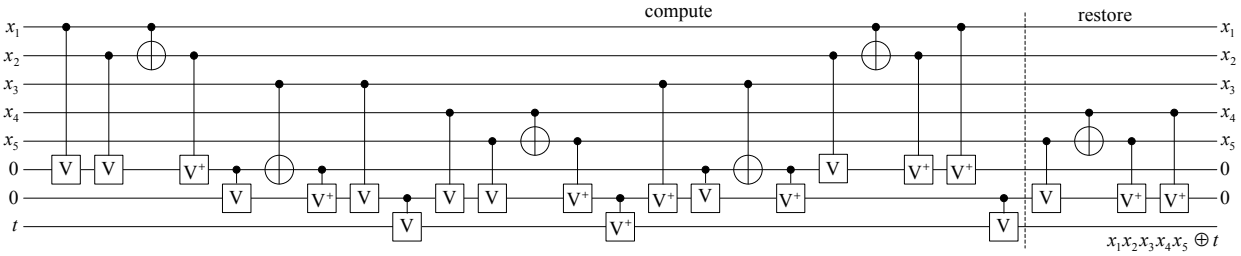


Fig. 11. Primitive quantum gate realization of six-qubit Toffoli gate.

to 46.2% AI improvement over [20]. As number of qubit is very limited in current quantum technologies, huge reduction of qubit with small increase of QC is desirable and as such our method makes a trade off between QC and AI. No result for $n > 15$ is reported in [20] and we can not compare those results. In general, we can say that our method of quantum primitive gate realizations of multiple-controlled Toffoli gates outperforms that of [20].

IV. Conclusion

Quantum algorithms are implemented using quantum circuits, which are designed using reversible circuit design techniques [1]. For this purpose multiple-controlled Toffoli gates are extensively used [3-9]. Beside quantum circuit design, multiple-controlled Toffoli gates are also extensively used for reversible designs of reversible and irreversible Boolean functions [10-19]. Therefore, efficient primitive quantum gate realizations of these gates are very important. Primitive quantum gate realization of three-qubit Toffoli gate using five primitive quantum gates is reported in [23]. However, a very limited attempt has been made for primitive quantum gate realizations of multiple-controlled (with more than two control lines) Toffoli gates. The best attempt so far has been reported in [20]. The method of [20] is unstructured and results of up to 15 control lines are reported.

In this paper, we present realization of three-qubit (two control lines) Toffoli gate using five primitive quantum gates, which is very similar to that of [23] with some rearrangement of the gates. We also present realization of four-qubit (three control lines) Toffoli gates using 13 primitive quantum gates and one ancilla input. We then present a general and recursive structure for realizations of $(n + 1)$-qubit ($n > 3$ control lines) Toffoli gates using $(\lceil n/2 \rceil + 1)$-qubit and $(\lfloor n/2 \rfloor + 1)$-qubit Toffoli gates with a trade off between number of primitive quantum gates (quantum cost) and number of ancilla inputs. Comparisons of quantum cost (QC) and ancilla inputs (AI) of our work with those of [20] show that for $n = 3$, we achieved 7.1% QC improvement over [20] without any further improvement of AI. For $n = 4$ to $n = 12$, we achieved 3.1% to 30.0% QC improvement and 25.0% to 50.0% AI improvement over [20]. For $n = 13$ to $n = 15$, our QC is 3.7% to 14.6% more than those of [20], but we achieved 36.4% to 46.2% AI improvement over [20]. As number of qubit is very limited in current quantum technologies, huge reduction of qubit with small increase of QC is desirable and as such our method makes a trade off between QC and AI. No result for $n > 15$ is reported in [20] and we can not compare those results. In general, our method of quantum primitive gate realizations of multiple-controlled Toffoli gates outperforms that of [20].

TABLE I. Quantum cost (QC) and ancilla input (AI) needed in primitive quantum gate realizations of multiple-controlled Toffoli gates

Control lines	Quantum cost of compute part	Quantum cost of restore part	Total quantum cost	Ancilla input
2	4	1	5	0
3	8	5	13	1
4	15	4	19	1
5	23	4	27	2
6	27	8	35	3
7	41	8	49	3
8	48	15	63	3
9	64	15	79	4
10	72	23	95	5
11	80	23	103	6
12	84	27	111	7
13	112	27	139	7
14	126	41	167	7
15	140	41	181	7
16	147	48	195	7
17	179	48	227	8
18	195	64	259	9
19	211	64	275	10
20	219	72	291	11

TABLE II. COMPARISON OF QUANTUM COST (QC) AND ANCILLA INPUT (AI) WITH THOSE OF [20]

Control lines	[20]		Our work		Improvement over [20]	
	QC	AI	QC	AI	QC (%)	AI (%)
3	14	1	13	1	7.1	0.0
4	26	2	19	1	26.9	50.0
5	38	3	27	2	28.9	33.3
6	50	4	35	3	30.0	25.0
7	62	5	49	3	21.0	40.0
8	74	6	63	3	14.9	50.0
9	86	7	79	4	8.1	42.9
10	98	8	95	5	3.1	37.5
11	110	9	103	6	6.4	33.3
12	122	10	111	7	9.0	30.0
13	134	11	139	7	-3.7	36.4
14	146	12	167	7	-14.4	41.7
15	158	13	181	7	-14.6	46.2

REFERENCES

[1] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information*. Cambridge University Press, 2000.

[2] D. Bacon and W. van Dam, "Recent progress in quantum algorithms," *Commun. ACM*, vol. 53, pp. 84-93, Feb. 2010.

[3] L. M. K. Vandersypen, M. Steffen, G. Breyta, C. S. Yannoni, R. Cleve, and I. L. Chuang, "Experimental realization of an order-finding algorithm with an NMR quantum computer," *Phys. Rev. Lett.*, vol. 85, pp. 5452-5455, Dec. 2000.

[4] L. M. K. Vandersypen, M. Steffen, G. Breyta, C. S. Yannoni, M. H. Sherwood, and I. L. Chuang, "Experimental realization of Shor's quantum factoring algorithm using nuclear magnetic resonance," *Nature*, vol. 414, no. 6866, pp. 883-887, 2001.

[5] N. Xu, J. Zhu, D. Lu, X. Zhou, X. Peng, and J. Du, "Quantum factorization of 143 on a dipolar-coupling nuclear magnetic resonance system," *Phys. Rev. Lett.*, vol. 108, p. 130501, Mar. 2012.

[6] I. L. Markov and M. Saeedi, "Constant-optimized quantum circuits for modular multiplication and exponentiation," *Quantum Info. Comput.*, vol. 12, no. 5-6, pp. 361-394, May 2012.

[7] A. Shafaei, M. Saeedi, and M. Pedram, "Reversible logic synthesis of k-input, m-output lookup tables," in *Design Automation and Test in Europe*, 2013 (to appear).

[8] A. M. Childs, R. Cleve, E. Deotto, E. Farhi, S. Gutmann, and D. A. Spielman, "Exponential algorithmic speedup by a quantum walk," in *ACM Symposium on Theory of Computing*, pp. 59-68, 2003.

[9] C.-F Chiang, D. Nagaj, and P. Wocjan, "Efficient circuits for quantum walks," *Quantum Info. Comput.*, vol. 10, no. 5, pp. 420-434, May 2010.

[10] D. Maslov, G. W. Dueck, and D. M. Miller, "Toffoli network synthesis with templates," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 24, no. 6, pp. 807-817, Jun. 2005.

[11] K. Fazel, M. Thornton, and J. Rice, "ESOP-based Toffoli gate cascade generation," in *IEEE Pacific Rim Conference on Communications, Computers, and Signal Processing*, pp. 206-209, Aug. 2007.

[12] D. Maslov, G. W. Dueck, D. M. Miller, and C. Negrevergne, "Quantum circuit simplification and level compaction," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 27, no. 3, pp. 436-444, Jun. 2008.

[13] J. E. Rice and V. Suen, "Using autocorrelation coefficient-based cost functions in ESOP-based Toffoli gate cascade generation," in *23rd Canadian Conference on Electrical and Computer Engineering (CCECE)*, May 2010.

[14] D. M. Miller, R. Wille, and R. Drechsler, "Reducing reversible circuit cost by adding lines," in *40th IEEE International Symposium on Multiple-Valued Logic (ISMVL)*, pp. 217-222, May 2010.

[15] Y. Sanaee and G. W. Dueck, "ESOP-based Toffoli network generation with transformations," in *40th IEEE International Symposium on Multiple-Valued Logic (ISMVL)*, pp. 276-281, May 2010.

[16] M. Arabzadeh , M. Saeedi, and M. Zamani, "Rule-based optimization of reversible circuits," in *Asia and South Pacific Design Automation Conference (ASP-DAC)*, pp. 849-854, 2010.

[17] N. M. Nayeem and J. E. Rice, "Improved ESOP-based synthesis of reversible logic," in *Reed-Muller Workshop*, May 2011.

[18] J. E. Rice and N. Nayeem, "Ordering techniques for ESOP-based Toffoli cascade generation," in *IEEE Pacific Rim Conference on Communications, Computers and Signal Processing (PacRim)*, pp. 274-279, Aug. 2011.

[19] N. M. Nayeem and J. E. Rice, "A shared-cube approach to ESOP-based synthesis of reversible logic," in *Facta Universitatis – Series: Electronics and Energetics*, vol. 24, no. 3, pp. 385-402, Dec. 2011.

[20] D. M. Miller, R. Wille, and Z. Sasanian, "Elementary quantum gate realizations for multiple-control Toffoli gates," in *41st IEEE International Symposium on Multiple-Valued Logic (ISMVL)*, May 2011.

[21] D. G. Cory, R. Laflamme, E. Knill, L. Viola, T. F. Havel, N. Boulant, G. Boutis, E. Fortunato, S. Lloyd, R. Martinez, C. Negrevergne, M. Pravia, Y. Sharf, G. Teklemariam, Y. S. Weinstein, and W. H. Zurek, *NMR Based Quantum Information Processing: Achievements and Prospects*, 2000. quant-ph/0004104.

[22] M. R. Geller, E. J. Pritchett, A. T. Sornborger, and F. K. Wilhelm, *Quantum Computing With superconductors I: Architectures*, 2006. quant-ph/0603224.

[23] A. Barenco, C. Bennett, R. Cleve, D. DiVinchenzo, M. Margolus, P. Shor, T. Sleator, J. Smolin, and H. Weinfurter, "Elementary gates for quantum computation," *Phys. Rev. A*, vol. 52, no. 5, pp. 3457-3467, Nov. 1995.

The Beneficial Effect of Filtering in Current Controlled Inverter

Md. Imran Azim[1], Md. Forhad Zaman[2], Mijanur Rahman[3]
[1, 2, 3] Department of Electrical & Electronic Engineering, Rajshahi University of Engineering & Technology (RUET), Bangladesh
imran.azim89@gmail.com, m.fzaman@yahoo.com, mijan.ieee@gmail.com

***Abstract*—An approach of harmonics component analysis, portraying the presence of certain amount of harmonics at the inverter output terminal, has been indicated in this paper. Some unwanted impacts are caused by the harmonics such as unbalance and excessive neutral currents, interference in nearby communication networks and disturbance to other consumer, torque pulsations in electric motor drives are and so on. Thus, the minimization of the harmonics contained in the output of a single phase current controlled inverter undergoing hysteresis modulation technique is important so as to get rid of these detrimental effects. With a view to reducing the assumed 3rd order harmonics, an LC low pass filter is used in this paper that blocks the harmonics and undeniably passes almost a sinusoidal output at the output terminal. It has been found from the simulation that the Total Harmonic Distortion (THD) is just above 33% during the existence of up to 3rd order harmonics, although it is predicted to be 0% in ideal case, and more importantly, that can be slumped to 0.00003% by the application of an LC low pass filter.**

***Index Terms*—Single Phase Inverter, Hysteresis Modulation, IGBT, Harmonics Impact, Fast Fourier Transform (FFT), Total Harmonic Distortion (THD), LC low pass Filter.**

I. INTRODUCTION

An inverter can be defined as dc to ac converter which symbolizes the change of dc voltage into a symmetric ac output voltage of desired magnitude and frequency as shown in "Fig. 1". The output voltage could be fixed or variable at a fixed or variable frequency. A variable output can be obtained by varying the input dc voltage and maintaining the gain of the inverter constant. The output waveforms of an ideal inverter should be sinusoidal. However, the waveforms of practical inverters are non-sinusoidal and contain certain harmonics which can be seen with ease in frequency domain. Due to the availability of high speed power semiconductor devices, the harmonic contents of output voltage can be minimized or reduced significantly by switching technique. IGBTs can be used as ideal switches to explain the power conversion techniques [1-2]. Some typical applications of inverter are variable speed ac drive; induction heating, uninterruptible power supplies (UPS), traction and HVDC [3].

Figure 1. General Block Diagram of Inverter.

Current Controlled Inverter is a special type of inverter, in which the output current can be controlled through hysteresis modulation techniques. Hysteresis current control is a kind of instantaneous feedback control model. The basic idea is to compare the given current signal, which is detected by the converter with actual inverter current signal [4].

Total Harmonic distortion (THD) is a measure of closeness in shape between a waveform and its fundamental component. It is given by the expression:

$$THD = \sqrt{\frac{\mathrm{Im}^2 - \mathrm{Im}_1{}^2}{\mathrm{Im}^2{}_1}} \tag{1}$$

Where, Im_1 is the RMS value of fundamental current component and Im is the RMS value of Current components. Therefore, it is needless to say that THD can be defined as the ratio of the RMS value of all odd number of non fundamental frequency terms to the RMS value of the fundamental frequency term. For the improvement of THD, a LC Low pass filter is appended at the output terminal that provides low harmonic impedance to ground [5].

This paper focuses on the advanced modulation technique, better known as hysteresis modulation and the introduction of suitable filtering in a proper way so that the harmonic components could be diminished to a permissible limit. It is certainly the case that this proposed methodology is going to aid any experiment involving pulse width modulated inverters.

II. SINGLE PHASE HYSTERESIS CURRENT CONTROLLED INVERTER

As the name suggests, the output current of the hysteresis inverter can be controlled by compelling the inverter line current to track a sinusoidal reference current within a specified error margin in order to attain an adequate switching optimization, excellent dynamic responses and high accuracy in steady-state operation regarded as a hysteresis modulation [6]. Line harmonic generation from this inverter depends principally on the particular switching pattern applied to the valves. The switching pattern of hysteresis inverters is produced through line current feedback and it is not pre-determined unlike the case, for instance, of Sinusoidal Pulse-Width Modulation (SPWM), where the inverter switching function is independent of the instantaneous line current and the inverter harmonics can be obtained from the switching

function harmonics [7]. The hysteresis modulation for power electronic converters is preferred for applications, where performance requirements are more demanding, such as to achieve good dynamic response, unconditional stability, and wide command-tracking bandwidth [8].

For simplicity, it is assumed that the dc voltage supplied by the DG source is divided into two not only constant but also balanced dc sources, as presented in the "Fig. 2", each of 230V. The RL element on the ac side represents the combined line and transformer inductance and losses. The inverter line current I, track a sinusoidal reference $i_{ref} = \sqrt{2}\, I_{ref} \sin(wt)$ Referring to "Fig. 2", when valve Q_1 is turned on, the inverter voltage is $v = v_C < v_{ref}$, this forces the line current, I to slope upward until the lower limit of the relay band $e_a(t) = -\varepsilon$ is reached. At that moment, the relay switches on Q_2 and the inverter voltage becomes $v = -v_C < v_{ref}$, this forces the line current, I to slope downward until the lower limit of the relay band is reached $e_a(t) = \varepsilon$

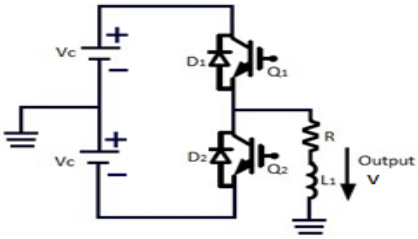


Figure 2. Single Phase Half Bridge Inverter.

The operation of the circuit illustrated in "Fig. 2", has been summarized in Table-I.

Table-I. Summarization of Circuit Operation

Q_1	Q_2	v
on	off	V_c
off	on	$-V_c$
on	on	0
off	off	0

Instantaneous inverter output voltage,

$$v = \sum_{n=1}^{\infty} 2\frac{2v_C}{n\pi}\sin(nwt) \tag{2}$$

Instantaneous inverter output current,

$$i = \sum_{n=1}^{\infty} 2\frac{2v_C}{n\pi\sqrt{R^2 + (nwL)^2}}\sin(nwt - \theta n) \tag{3}$$

Where, θn is the lagging angle of the inductor.

Modulated current is the difference between the reference current and the inverter output current. If the actual current is more than the given value, then it is decreased by changing the switching state and vice versa. So, the actual current changes around the reference current waveform and hysteresis current control makes the deviation within a certain range [9].

$$e_a(t) = i_{ref} - i \tag{4}$$

III. Simulation of the Hysteresis Inverter

For making the simulation possible, necessary parameters like supply voltage=230 V, frequency=50 Hz, load resistor=15 Ohm and load inductor =5 H are considered deliberately with assuming that there is no harmonics prevalent at the output, which means output response is forecasted to be sinusoidal.

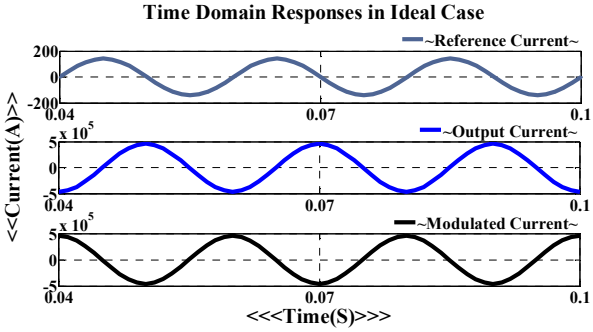


Figure 3. Time Domain Responses without Harmonics.

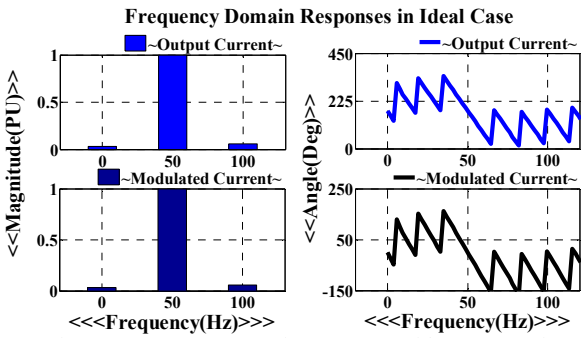


Figure 4. Frequency Domain Responses without Harmonics.

According to the depiction, "Fig. 3" and "Fig. 4", deal with the time domain and frequency domain responses of the inverter output current and the modulated current, implying the difference between reference and inverter current while there is no existence of harmonics. Here, the obtained THD is 0%. But in practical situations, it is cumbersome to get this. Now, 3rd harmonics component is considered so that the performance of hysteresis current controlled inverter under harmonic prevalent state can be realized.

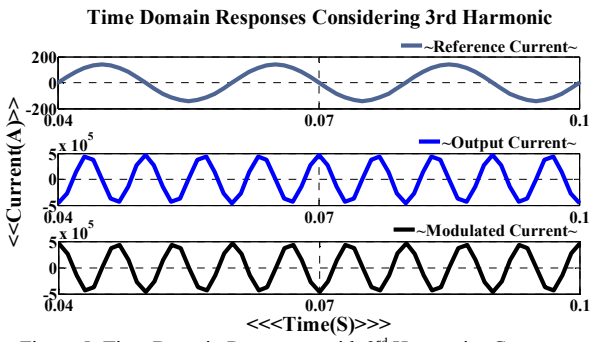


Figure 5. Time Domain Responses with 3rd Harmonics Component.

"Figure 5", reveals the time domain responses hysteresis current controlled inverter when 3rd order harmonics is present

at the output terminal. As is observed, none of these figures show pure sinusoidal responses due to the existence of unwanted frequency called harmonics.On the other hand, "Fig. 6", represents the responses in frequency domain. It is clearly seen that others frequency components are there with the fundamental frequency component, symbolizing the undesised term regarded as 3rd order harmonics. Using "Eq. 1", under this circumstance, the THD is evaluated and is found 33.33% which is not satisfactory at all. There is no denial that for better performance it is important to evade this 3rd harmonics.

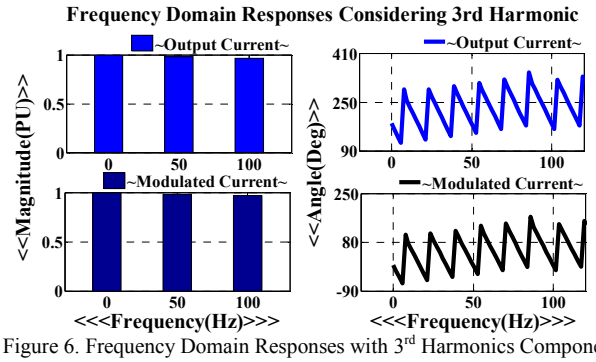


Figure 6. Frequency Domain Responses with 3rd Harmonics Component.

IV. Application of an LC Low Pass Filter

It is worthy to mention that the basic hysteresis technique is affected by the drawbacks of a variable switching frequency, large ripple current in steady state, generation of sub harmonic components and so on [10] that urges the necessity of filtering the output response [11]. The implementation of an LC filter, having L=2H and C=1000F at the inverter ac terminals could trigger a parallel resonance, which tends to amplify the harmonic components in ac network, owing to the fact that the filter capacitance has a profound impact on the performance of harmonics and so the harmonics can be brought into a lower state [12].

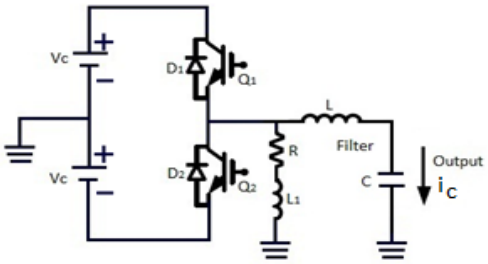


Figure 7. Single Phase Half Bridge Inverter with Filter.

Now, output is taken across the capacitor with 1000F capacitance. In this case the inverter output current becomes,

$$i_C = \sum_{n=1}^{\infty} 2\frac{2v_C}{n\pi\sqrt{\left(-1/nwc\right)^2}}\sin(nwt-\theta c) \qquad (5)$$

Where, θc is the current leading angle of the capacitor.

V. Simulation of the Inverter with Filter

Finally, in order to visualize the noteworthy performance of the filter, having leading current capability, the output response derived from "Eq. 5" as well as the modulated current picked out across the condenser, as seen in "Fig. 7", are simulated.

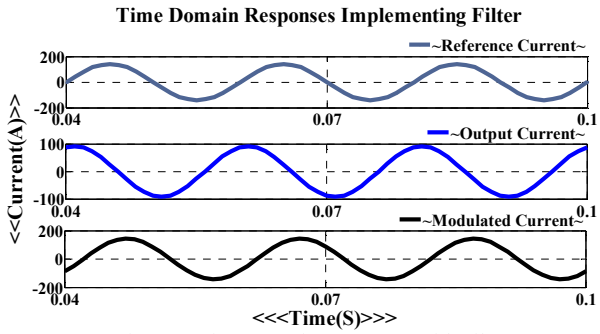


Figure 8. Time Domain Responses with Filter.

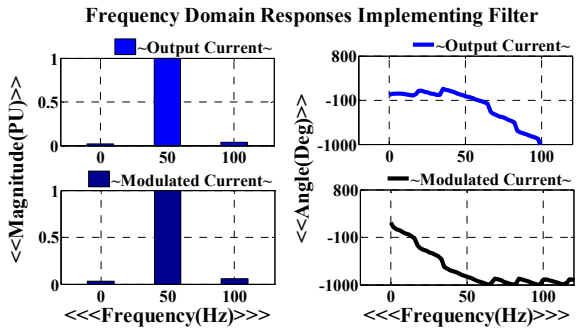


Figure 9. Frequency Domain Responses with Filter.

As far as "Fig. 8", and "Fig. 9", are concerned, as is expected, not just desired sinusoidal time domain inverter output response is obtained but also the received modulated current response is sinusoidal and in each case, it is noticed that the fundamental component has the highest amplitude. Lastly, "Eq. 1" is once again utilized and the calculated THD is 0.00003% that can be regarded as a negligible amount. Thence, harmonics have been suppressed to a great extent.

VI. Conclusion

When ideal response has been estimated, Total Harmonic Distortion (THD) is 0%, which is not plausible in real life utilities. At normal condition, harmonics have to be taken into consideration and if 3rd order harmonic component is considered then there exists about 33.33% THD that is undisputedly beyond the satisfaction. But as soon as an LC low pass filter is appended, it has been dropped to approximately 0.00003%, notifying a vast improvement due to the effect of capacitance on the controlling of harmonics. It has been found from experimental results [13] that the THD of three level, five level, seven level and nine level inverter systems are 10.95%, 4.82%, 4.65% and 3.09% respectively, referring that the THD is declining with the enhancement of the level of inverter. In contrast, only 0.00003% THD is derived in case of single phase inverter mentioned in this

paper. Therefore, a remarkable improvement has been achieved, when it comes to compare with respect to THD.

In future, this strategy may find an extensive utility in smart grid network integrated with solar energy, in which the derived solar dc voltage is needed to convert into ac in a convenient way, where harmonic free operation is obligatory to be ensured. In addition, in order to get three phase ac voltages, harmonics control in three phase inverters using this methodology would be interesting in upcoming days.

REFERENCES

[1] Md. Imran Azim, Md. Forhad Zaman, Tamal Sarkar, "Better Performance Obtained in Hysteresis Current Controlled Inverter By Filter Implementation," ICAEE, Dhaka, December, 2013, unpublished.

[2] M. H. Rashid, Power Electronics Circuits, Devices, and Applications, Third Edition, Dorling Kindersley Pvt. Ltd, 2006, Chapter-6.

[3] I. A, H. Rahman, "Analysis & Simulation of Total Harmonic Distortion in Hysteresis Inverter," International Journal of Advanced Research in Computer Engineering & Technology (IJARCET). vol. 2, issue. 3, March, 2013, pp. 574-578.

[4] Wang Zhao-an, HuangJun, Power Electronic Technology [M].Beijing China Machine Press 2000, pp. 47-50.

[5] A. J. Onah, "Harmonics: Generation and Suppression in AC System Networks," Nigerian Journal of Technology, vol. 31, no. 3, November, 2012, pp. 293-299.

[6] M. Mahalakshmi, V. Kumaresan, "Hysteresis Modulation of Cascaded Inverter". International Journal of Emerging trends in Engineering and Development, vol. 3, issue. 2, April, 2012.

[7] A. Albanna, Modeling and Simulation of Hysteresis Current Controlled Inverters Using MATLAB, In Tech, 2011, pp. 98-120.

[8] M. P. Kazmierkowski, L. Malesani, "Current control techniques for three-phase voltage-source PWM converters: A survey," IEEE Trans. Ind. Electron., vol. 45, no. 5, October, 1998, pp.691-703.

[9] Y. L. Wang, Q. L. Guo, "Hysteresis Current Control technique based on Space Vector Modulation for Active Power Filter," International Journal of Power Electronics and Drive System, vol.1, no.1, September, 2011, pp. 1-6.

[10] L. Malesani, P. Mattavelliand, P. Tomasin, "High-Performance Hysteresis Modulation Technique for Active Filters," IEEE Trans. Power Electronics, vol. 12, no. 5, September, 1997.

[11] I. A, H. Rahman. "Harmonics Reduction of a Single Phase Half Bridge Inverter," Global Journal of Research in Engineering, vol.13. issue.4, version 1, 2013, pp. 9-12.

[12] B. Singh, K. Al-Haddad, A. Chandra. "A Review of Active Filters for Power Quality Improvement," IEEE Transactions on Industrial Electronics, vol. 46, no. 5, 1999, pp. 960-971.

[13] G. Mahesh, M. Kumar, S. R. Reddy, "Simulation and Experimental Results of 7-Level Inverter System," Research Journal of Applied Sciences, Engineering and Technology. vol. 3, no. 2, 2011. pp. 88-95.

[14] I. A, H. Rahman, "Minimization of Total Harmonic Distortion of a Single Phase Inverter," American Academic & Scholarly Research Journal, vol.5, no.4, May, 2013, pp.193-199.

[15] I. A, H. Rahman. "Simulation of Load variation Impacts on a Delta Modulated Inverter," Global Journal of Research in Engineering, vol.13, issue .6, version 1, 2013, pp. 23-27.

[16] D. E. Steeper and R.P. Stratford. "Reactive Compensation and Harmonic Suppression for Industrial Power Systems Using Thyristor Converters," IEEE Trans. Industry Applications, vol. 12, 1976, pp. 232-254.

[17] Md. I. A, Dr. Md. F. Rahman, "Performance Improvement of Phase Displacement Modulated Inverter," International Journal of Electrical Energy, vol. 1, no. 4, December, 2013, pp.300-303.

Irregular total labellings of Ladder

K M Mominul Haque
Department of Computer Science and Engineering
Sylhet International University
Sylhet, Bangladesh
momin66@gmail.com

Umme Narseen Khanam
Department Computer Science
Jaintapur Tayob Ali Technical College
Jaintapur, Sylhet, Bangladesh

***Abstract*—The total edge irregularity strength *tes(G)* and total vertex irregularity strength *tvs(G)* are invariants analogous to irregular strength *s(G)* of a graph *G* for total labellings. Bača et al. [Discrete Mathematics, 307: 1378-1388, (2007)] determined the bounds and precise values for some families of graphs concerning these parameters. In this paper, we show the exact values of the total edge irregularity strength $tes(L_n) = n$ total vertex irregularity strength and $tvs(L_n) = \lfloor \frac{n}{2} \rfloor + 1$ for the Ladder L_n .**

***Index Terms*— Irregular total labelling; Ladder Graphs; Total labeling.**

I. Introduction

We consider only finite undirected graphs without loops or multiple edges. Let $G = (V, E)$ be a graph with vertex set V and edge set E.

An edge irregular total k-labelling of a graph G is a labelling of the vertices and edges with labels $1,\ldots, k$ such that for every two different edges their weights are distinct where the weight of an edge is the sum of its label and the labels of its two end vertices. A vertex irregular total k-labelling of a graph G is a labelling of the vertices and edges with labels $1,\ldots, k$ such that for every two different vertices their weights are distinct where the weight of a vertex is the sum of its label and the labels of its incident edges. The minimum k for which the graph G has an edge irregular total k-labelling is called the total edge irregularity strength of the graph G, *tes(G)*. Analogously, the minimum k for which there exists a vertex irregular total k-labelling is called the total vertex irregularity strength of G, *tvs(G)*.

The notions of the total edge irregularity strength and total vertex irregularity strength were first introduced by Bača et al. [1]. They may be taken as an extension of the irregularity strength of a graph [2, 4, 6, 8, 9, 12, and 13]. In [1], the authors put forward the lower bounds of *tes(G)* and *tvs(G)* in terms of the maximum degree Δ, minimum degree δ, $|E(G)|$ and $|V(G)|$, which may be stated as the Theorems 1.1 and 1.2:

Theorem 1.1. $tes(G) \geq \max\left\{\left\lceil \frac{\Delta+1}{2} \right\rceil, \left\lceil \frac{|E(G)|+2}{3} \right\rceil\right\}$.

Theorem 1.2. $tvs(G) \geq \left\lceil \frac{|V(G)|+\delta}{\Delta+1} \right\rceil$.

Bača et al. [1] then determined the exact values of the total edge irregularity strength for path P_n, star S_n, wheel W_n and friendship graph F_n, and obtained the exact values of the total vertex irregularity strength for star S_n, complete graphs K_n, cycle C_n and prism D_n. The author proved that irregular total labellings of Generalized Petersen graphs *P(n, k)* and Irregular Total Labelling of Möbius Ladder in [10, 11]. We refer the readers for some recent results [3, 5, and 7]. In this paper, we deal with the Ladder Graphs.

The ladder L_n is defined to be a graph on $2n$ vertices with $V(L_n) = \{u_i, v_i : 0 \leq i \leq n-1\}$ and $E(L_n) = \{u_i\ v_i : 0 \leq i \leq n-1\} \cup \{u_i u_{i+1}, v_i v_{i+1} : 0 \leq i \leq n-1\}$.

In Figure 1.1, we show the Ladder L_5.

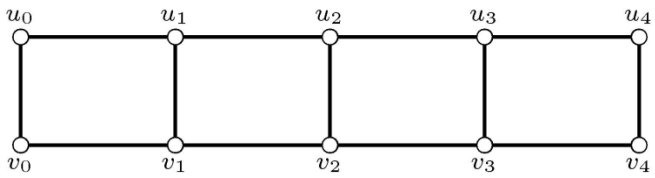


Figure 1.1. Ladder graph for L_5

II. Main Results

2.1 Irregular total labelling of L_n

Theorem 2.2.1. $tes(L_n) = n$.

Proof. For $0 \leq i \leq n-1$, we construct the function f as follows:

$$\begin{aligned} f(u_i) &= 1, & 0 \leq i \leq n-1, \\ f(v_i) &= n, & 0 \leq i \leq n-1, \\ f(u_i v_i) &= i+1, & 0 \leq i \leq n-1, \\ f(u_i u_{i+1}) &= i+1, & 0 \leq i \leq n-2, \\ f(v_i v_{i+1}) &= i+2, & 0 \leq i \leq n-2. \end{aligned}$$

Since

$$\begin{aligned} wt(u_i u_{i+1}) &= f(u_i) + f(u_i u_{i+1}) + f(u_{i+1}) \\ &= 1 + (i+1) + 1 \\ &= 3 + 1, \\ wt(u_i v_i) &= f(u_i) + f(u_i v_i) + f(v_i) \\ &= 1 + (i+1) + n \\ &= n + 2 + i, \end{aligned}$$

$$\begin{aligned} wt(v_i v_{i+1}) &= f(v_i) + f(v_i v_{i+1}) + f(v_{i+1}) \\ &= n + (i+2) + n \\ &= 2n + 2 + i. \end{aligned}$$

the weights of edges of L_n under the labeling f constitute the set $\{3, 4, \cdots, 3n\}$and the function f is a map from $V(L_n) \cup E(L_n)$ into $\{1, 2, \cdots, n\}$.

Clearly, that the total labelings f have the required properties of an edge irregular total labeling. We then have $tvs(L_n) \leq n$· By Theorem 1.1, $tvs(L_n) \geq \lceil \frac{|E(G)|+2}{3} \rceil = \left\lceil \frac{3n}{3} \right\rceil = n$. This concludes the proof.

In Figure 2.2.1, we show the edge irregular total labeling for L_{10}.

Figure

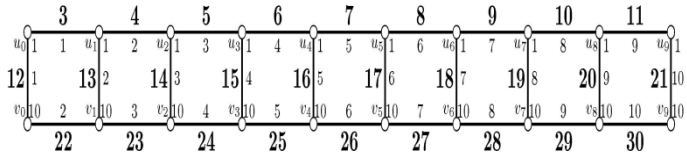


Figure 2.2.1. Edge irregular total labellings for L_{10}

Theorem 2. 2. 2. $tvs(L_n) = \left\lfloor \frac{n}{2} \right\rfloor + 1$.

Proof.

For$0 \leq i \leq n-1$, we construct the function f as follows:

Case 1. $n \bmod 2 = 0$.

$$f(u_i u_{i+1}) = \begin{cases} i+1, & 0 \leq i \leq \frac{n}{2} - 1, \\ n - i, & \frac{n}{2} \leq i \leq n-1, \end{cases}$$

$$f(v_i v_{i+1}) = \begin{cases} i+1, & 0 \leq i \leq \frac{n}{2} - 2, \\ n - i, & \frac{n}{2} - 1 \leq i \leq n-1, \end{cases}$$

$$f(u_i v_i) = \begin{cases} 1, & i = 0, \\ i, & 1 \leq i \leq \frac{n}{2}, \\ n - i, & \frac{n}{2} + 1 \leq i \leq n-1, \end{cases}$$

$$f(u_i) = \begin{cases} 1, & i = 0, \\ i+2, & 1 \leq i \leq \frac{n}{2} - 1, \\ \frac{n}{2} + 1, & i = \frac{n}{2}, \\ n - i, & \frac{n}{2} + 1 \leq i \leq n-2, \\ 2, & i = n-1, \end{cases}$$

$$f(v_i) = \begin{cases} 2, & i = 0, \\ i+3, & 1 \leq i \leq \frac{n}{2} - 2, \\ \frac{n}{2} + 1, & i = \frac{n}{2} - 1, \\ n - i + 1, & \frac{n}{2} \leq i \leq n-2, \\ 3, & i = n-1. \end{cases}$$

Since

$$wt(u_i) = \begin{cases} 1 + (i+1) + 1 = i + 3, & i = 0, \\ ((i-1) + 1 + i + (i+1) + (i+2) \\ = 4i + 3, & 1 \leq i \leq \frac{n}{2} - 1, \\ \big((i-1) + 1\big) + i + (n-i) + \left(\frac{n}{2} + 1\right) \\ = 2n + 1, & i = \frac{n}{2}, \\ \big(n - (i-1)\big) + (n-i) + (n-i) + (n-i) \\ = 4n - 4i + 1, & \frac{n}{2} + 1 \leq i \leq n-2, \\ \big(n - (i-1)\big) + (n-i) + 2 \\ = 2n - 2i + 3, & i = n-1. \end{cases}$$

$$wt(v_i) = \begin{cases} \big(n - (n-1)\big) + (i+1) + 2 \\ = i + 4, & i = 0, \\ \big((i-1) + 1\big) + i + (i+1) + (i+3) \\ = 4i + 4, & 1 \leq i \leq \frac{n}{2} - 2, \\ \big((i-1) + 1\big) + i + (n-i) + \left(\frac{n}{2} + 1\right) \\ = 2n, & i = \frac{n}{2} - 1, \\ \big(n - (i-1)\big) + i + (n-i) + (n-i+1) \\ = 2n + 2, & i = \frac{n}{2}, \\ \big(n - (i-1)\big) + (n-i) + (n-i) + (n-i+1) \\ = 4n - 4i + 2, & \frac{n}{2} + 1 \leq i \leq n-2, \\ \big(n - (\ i-1)\big) + (n-i) + 3 \\ = 2n - 2i + 4, & i = n-1. \end{cases}$$

The weights of vertices of L_n under the labeling f constitute the set $\{3, 4, \cdots, 2n+2\}$ and the function f is a map from $V(L_n) \cup E(L_n)$ into $\{1, 2, \cdots, \frac{n}{2} + 1\}$ for even n.

Case 2. n mod 2 = 1.

$$f(u_i u_{i+1}) = \begin{cases} i+1, & 0 \leq i \leq \lfloor \frac{n}{2} \rfloor - 1, \\ n - i, & \lfloor \frac{n}{2} \rfloor \leq i \leq n-1, \end{cases}$$

$$f(v_i v_{i+1}) = \begin{cases} i+1, & 0 \leq i \leq \left\lfloor \frac{n}{2} \right\rfloor - 2, \\ \left\lfloor \frac{n}{2} \right\rfloor + 1, & i = \left\lfloor \frac{n}{2} \right\rfloor - 1, \\ n - 1, & \left\lfloor \frac{n}{2} \right\rfloor \leq i \leq n-1, \end{cases}$$

$$f(u_i v_i) = \begin{cases} 1, & i = 0, \\ i, & 1 \le i \le \left\lfloor \frac{n}{2} \right\rfloor - 1, \\ n - i, & \lfloor \frac{n}{2} \rfloor \le i \le n - 1, \end{cases}$$

$$f(u_i) = \begin{cases} 1, & i = 0, \\ i + 2, & 1 \le i \le \lfloor \frac{n}{2} \rfloor - 1, \\ \lfloor \frac{n}{2} \rfloor + 1, & i = \lfloor \frac{n}{2} \rfloor, \\ n - i, & \lfloor \frac{n}{2} \rfloor + 1 \le i \le n - 2, \\ 2, & i = n - 1, \end{cases}$$

$$f(v_i) = \begin{cases} 2, & i = 0, \\ i + 3, & 1 \le i \le \lfloor \frac{n}{2} \rfloor - 2, \\ \left\lfloor \frac{n}{2} \right\rfloor + 1, & i = \left\lfloor \frac{n}{2} \right\rfloor - 1 \ or \ i = \lfloor \frac{n}{2} \rfloor, \\ n - i + 1, & \left\lfloor \frac{n}{2} \right\rfloor + 1 \le i \le n - 2, \\ 3, & i = n - 1. \end{cases}$$

Since

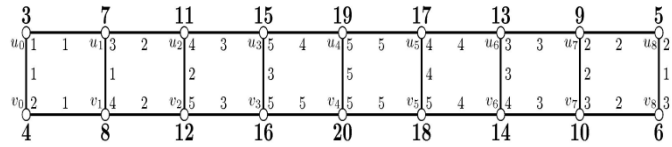
$$wt(u_i) = \begin{cases} (n - (n-1) + (i+1) + 1 \\ = i + 3, & i = 0, \\ ((i-1)+1) + i + (i+1) + (i+2) \\ = 4i + 3, & 1 \le i \le \lfloor \frac{n}{2} \rfloor - 1, \\ \left((i-1)+1\right) + (n-i) + (i+1) + \left(\lfloor \frac{n}{2} \rfloor + 1\right) \\ = 2n + 1, & i = \left\lfloor \frac{n}{2} \right\rfloor, \\ (n - i + 1) + (n - i) + (n - i) + (n - i) \\ = 4n - 4i + 1, & \lfloor \frac{n}{2} \rfloor + 1 \le i \le n - 2, \\ (n - i + 1) + (n - i) + 2 \\ = 2n - 2i + 3, & i = n - 1. \end{cases}$$

$$wt(v_i) = \begin{cases} 1 + (i+1) + 2 \\ = i + 4, & i = 0, \\ \left((i-1)+1\right) + i + (i+1) + (i+3) \\ = 4i + 4, & 1 \le i \le \lfloor \frac{n}{2} \rfloor - 2, \\ \left((i-1)+1\right) + i + \left(\lfloor \frac{n}{2} \rfloor + 1\right) + \left(\lfloor \frac{n}{2} \rfloor + 1\right) \\ = 2n - 2, & i = \lfloor \frac{n}{2} \rfloor - 1, \\ \left(\left\lfloor \frac{n}{2} \right\rfloor + 1\right) + (n - i) + \left(n - i) + (\lfloor \frac{n}{2} \rfloor + 1\right) \\ = 2n + 2, & i = \lfloor \frac{n}{2} \rfloor, \\ \left(n - (i-1)\right) + (n - i) + (n - i) + (n - i + 1) \\ = 4n - 4i + 2, & \lfloor \frac{n}{2} \rfloor + 1 \le i \le n - 2, \\ \left(n - (\ i - 1)\right) + (n - i) + 3 \\ = 2n - 2i + 4, & i = n - 1. \end{cases}$$

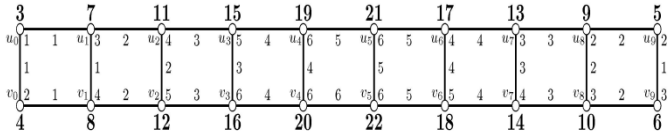
The weights of vertices of L_n under the labeling *f* constitute the set $\{3, 4, \cdots, 2n + 2\}$ and the function *f* is a map from $V(L_n) \cup E(L_n)$ into $\{1, 2, \cdots, \lfloor \frac{n}{2} \rfloor + 1\}$ for odd *n*.

From case 1 and case 2, the total labeling *f* has the required properties of vertex irregular total labeling. We then have $tvs(L_n) \le \lfloor \frac{n}{2} \rfloor + 1$. However, by Theorem 1.2, $tvs(G) \ge \lceil \frac{|V(G)|+\delta}{\Delta+1} \rceil = \lceil \frac{2n+2}{4} \rceil = \lfloor \frac{n}{2} \rfloor + 1$. This concludes the proof.

In Figure 2.2.2, we show the vertex irregular total labellings for L_9 and L_{10}.



(L_9)



(L_{10})

Figure 2.2.2. Vertex irregular total labellings for L_9 and L_{10}

III. Conclusion

In this paper, we apply the Bača's determination the bounds and precise values for some families of graphs concerning of parameters the total edge irregularity strength *tes(G)* and total vertex irregularity strength *tvs(G)*. We hope that the research along this direction can be continued, and in fact some results in this paper have already constituted a platform for further discussion concerning other graphs.

IV. Acknowledgment

The author is deeply indebted to the editor and reviewer for careful and thoughtful reading of the manuscript and their useful suggestions and comments which helped to improved the presentation of results.

References

[1] M. Bača, S. Jendrol, M. Miller, J. Ryan, "On irregular total labellings", *Discrete Mathematics*, 307(2007), 1378–1388.

[2] J. L. Baril, H. Kheddouci and O. Togni, "The irregularity strength of circulant graphs", *Discrete Mathematics*, 304(2005), 1–10.

[3] S. Brandt, J. Mišskuf, D. Rautenbach, "On a conjecture about edge irregular total labeling", *Journal of Graph Theory*, in press.

[4] T. Bohman and D. Kravitz, "On the irregularity strength of trees", *Journal of Graph Theory*, 45(4)(2004), 241–254.

[5] J. Ivančo and S. Jendrol, "The total edge irregularity strength of trees", *Discussiones Mathematicae. Graph Theory*, 26(3)(2006), 449–456.

[6] S. Jendrol, M. TkáčThe "Irregularity strength of tK_p", *Discrete Mathematics*, 145(1995), 301–305.

[7] S. Jendrol, J. Miškuf and R. Soták, "Total edge irregularity strength of complete graphs and complete bipartite graphs", *Electronic Notes in Discrete Mathematics*, 28(2007), 281–285.

[8] S. Jendrol and V. ˇZoldák, "The irregularity strength of generalized Petersen graphs", *Mathematica Slovaca*, 45(2)(1995), 107–113.

[9] A. Meissner, M. A. Niwińska and K. T. Zwierzynski, "Computing an irregularity strength of selected graphs", *Electronic Notes in Discrete Mathematics*, 24(2006), 137–144.

[10] Khandoker Mohammed Mominul Haque, "Irregular Total Labellings of Generalized Petersen graphs", *Theory of Computing Systems*, 50(2012), 537–544.

[11] Kh. Md. Mominul Haque, "Irregular Total Labellings of Möbius Ladder", *SUST Studies*, to appear.

[12] O. Togni, "Irregularity strength of the toroidal grid", *Graphs and combinatorics*, 165/l66 (1997), 609–620.

[13] O. Togni, "Irregularity strength and compound graphs", *Discrete Mathematics*, 218(2000), 235–243.

A compact W-shaped 2.45 GHz RFID Tag Antenna Design for UHF RFID Applications

Mukit Sarkar
Department of Electronics & Communication Engineering
Khulna University of Engineering & Technology, Khulna-9203, Bangladesh, E-mail: aronob_ece05@yahoo.com
Md. Abul Hossain
Department of Electronics & Communication Engineering
Khulna University of Engineering & Technology, Khulna-9203, Bangladesh, E-mail: hossain_ece24@yahoo.com
Md. Reshad ul Hoque
Department of Electronics & Communication Engineering
Khulna University of Engineering & Technology, Khulna-9203, Bangladesh, E-mail: shuvo12_reshad@yahoo.com
Md. Shahabul Alam
Department of Electrical & Electronics Engineering
Khulna University of Engineering & Technology, Khulna-9203, Bangladesh, E-mail: alam_shahabul@yahoo.com

***Abstract*—This paper represents design, analysis & investigation of an radiator for UHF RFID applications. This antenna covered the frequency spectrum 2392.86-2457.14MHz with bandwidth 14.28MHz which is sufficient for RFID tag applications. We have designed as well as simulated this antenna using Numerical Electromagnetic Code (NEC) antenna modeling software package which is based on the method of moments solution and after simulation we have got SWR exactly 1.097 having reflection coefficient -26.68 without filtering and after passing through a low-pass L-network the SWR value is sharply 1.00054 with reflection coefficient -71.386 at 2.45 GHz band. The total antenna gain is excellent and it is approximately 22.3 dBi. We have achieved the decent impedance that is 48.36-j4.95 ohm before filtering and 49.99-j0.029 ohm after filtering. Afterward we have fabricated this designed antenna on a low-cost printed circuit board with ground plane including dielectric constant 4.4 & thickness of the metal 1.58mm. The efficiency of the proposed antenna is 99.96%. The measurement shows that the SWR value is 1.05404 at 0 dB attenuation.**

Keywords—Electrically Small Antenna (ESA); Low-pass L-network; Numerical Electromagnetic Code (NEC); RFID Tag Antenna; Standing Wave Ratio (SWR);

I. INTRODUCTION

The application of RFID is increasing day by day in retail industry & government due to tremendous technology developments in integrated circuits & radios. Recently enormous research activities have been focused on the Radio Frequency Identification (RFID) systems for identification & tracking of individuals or social networks due to their massive applications in bio-engineering, security identifications, access control, tolling, transport application, e-passport, container logistic item, distribution retail etc. The RFID system was used for identification of aircraft during 2nd World War. The operating frequency bands for RFID applications are namely 1.25 KHz, 134.2 KHz, 13.56 MHz, 400 MHz & 865-869 MHz for Europe, 902-928 MHz for USA & 952-954 MHz for Japan. Aside these there have some operating bands in microwave applications such as 2.45 GHz & 5.8 GHz [1]-[2]. A miniature hybrid loop antenna has been proposed by S. -Yuan Chen in which the peak antenna gain measured at 2.45GHz is only 2.2 dBi which is much more limited than our proposed antenna for RFID applications [3]. A small on-chip-antenna (OCA) with a RFID tag at 2.45GHz circuit with an inductive coupling model and it is designed based on a CMOS technology with chip area is 1 X 0.5 mm2 [4] whereas another OCA with a RFID tag [6] on a chip area 0.64 X 0.64 mm2 at 2.45GHz for near field analysis. The inductively coupled RFID antennas namely arc-shape and dual-body configurations was designed [5] with some tag chips whose directivities 1.99dBi and 5.62 dBi at 915MHz. A miniaturized 2.45GHz RFID tag antenna designed on 0.5mm thick FR4 substrate with area 33.55 mm X 8.54mm [7] and obtained peak gain value of -0.2 dBi and the radiation efficiency 56%. The Finite Difference Time Domain (FDTD) 3D electromagnetic simulation method was used to simulate the desired results and optimized the antenna structure. A compact E-shaped RFID tag antenna designed for UHF RFID transponder [8]. The operating band of this antenna is 914MHz having dielectric constant 4.4 and thickness of the substrate is 1.58mm. The antenna has bandwidth 2MHz with the SWR value at 0 dB attenuation is 1.068.

II. ANTENA DESIGN AND THEORY AND CONFIGURATIO

The choice of antenna physical dimensions is driven by more convenience and more compact with its sufficient gain, efficiency and bandwidth. Wheeler defined an electrically small antenna (ESA) as one whose maximum dimension is less than $\frac{\lambda}{2\pi}$ [9]. This relation is usually expressed as:

$$Ka < 1$$

Where, $k = 2\pi/\lambda$ (radians/meter), λ=Free space wavelength (meters) and a=Radius of sphere enclosing the maximum dimension of the antenna (meters).

An electrically small antenna contained an inherent minimum value of Q. This places a limit on the attainable impedance bandwidth of an Electrically Small Antenna (ESA).The higher the antenna Q the smaller the impedance bandwidth. The minimum Q for an electrically small linear antenna in free space is expressed as:

$$Q_L = \frac{1}{k^3 a^3} + \frac{1}{ka} \quad (1)$$

The minimum Q for an ESA which is circularly polarized is:

$$Q_{cp} = \frac{1}{2}\left[\frac{1}{k^3 a^3} + \frac{1}{ka}\right] \quad (2)$$

Equations (1) and (2) assume a perfect lossless matching network. The approximate bandwidth for an RLC type circuit in terms of Q is:

$$\mathrm{BW} = \frac{s-1}{Q\sqrt{s}} \quad (3)$$

Where S = S : 1 VSWR

BW = Normalized Bandwidth.

In this research activity we have designed a compact line antenna that looks like as English capital letter W which is shown in Fig. 1. The antenna have wire radius 0.5mm & input voltage 5V. The antenna has the length 63mm and width 25.3mm. The miniaturized structure of all radiating elements has been presented on the top of the substrate in the Figure 1. The antenna has printed on a FR4 substrate whose dielectric constant 4.4. The reverse face of the substrate contains the copper sheet that acts as a ground plate and for this plate the side beams are reflected back towards main beam which heightened the overall directivity of the antenna to appropriate level. The excitation is given directly in the arm number 9 by 50 ohm coaxial SMA which is mentioned by a small circle. When the feeding point is given in this position then the VSWR of the overall structure diminished significantly. Aside this a series low pass L- network with XL=11.6 nH and Xc =2.13pF is used for proper impedance matching and best consequence of gain, VSWR, return loss, efficiency & overall performance. We have got optimized antenna performance after excited directly and the resonance found at 2.45 GHz when the whole structure of proposed antenna meets the desired dimensions with capacitive as well as inductive reactance. This reactance affects the great impact on antenna input impedance for decent performance which is highly appreciated. Here the straight arms of W-shaped antenna have made a resonance at a lower frequency with a narrow bandwidth.

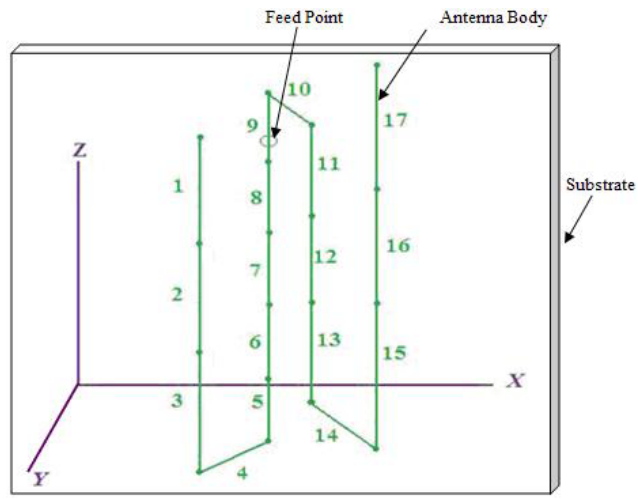


Fig. 1. Structural Geometry and dimensions of W-Shaped Antenna

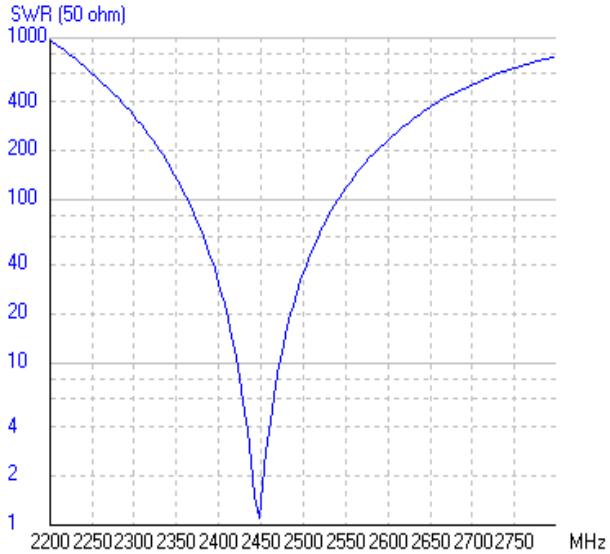
The total antenna dimensions are given in Table 1. The antenna length varies from 5.94mm to 23.2mm.

TABLE I. DIMENSION OF THE PROPOSED ANTENNA

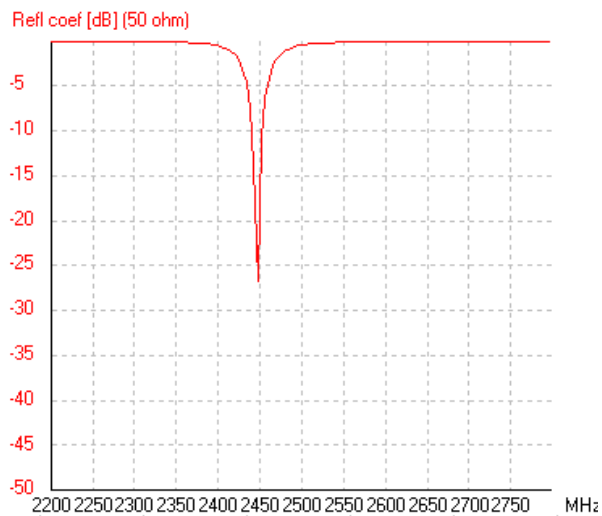
Wire No.	Dimension (mm)	Wire No.	Dimension (mm)
1	17.5	10	9.39
2	16	11	13
3	23.2	12	12
4	5.94	13	17.9
5	8	14	9.97
6	11	15	19
7	13.2	16	21
8	11.8	17	23
9	12.9		

III. SIMULATION AND PRACTICAL MEASUREMENTS

The Proposed W-shape antenna has been designed & analysis by a popular antenna modeling software package NEC which is based on the method of moments for the solution of the electric as well as magnetic field integral equations for thin wire & surface antennas. At first we simulated the overall performance of this antenna in the absence of low pass L-network. The SWR & reflection coefficient is shown in Figure 2. We found the SWR exactly 1.097 and reflection coefficient is -26.68 at nearly 2.45 GHz frequency response. When we used low pass L-network (XL=nH, Xc=pF) then got better result that is SWR is exactly 1.00054 which is very approximately 1 and reflection coefficient -71.386 at frequency response closely 2.45GHz as shown in Figure 3.

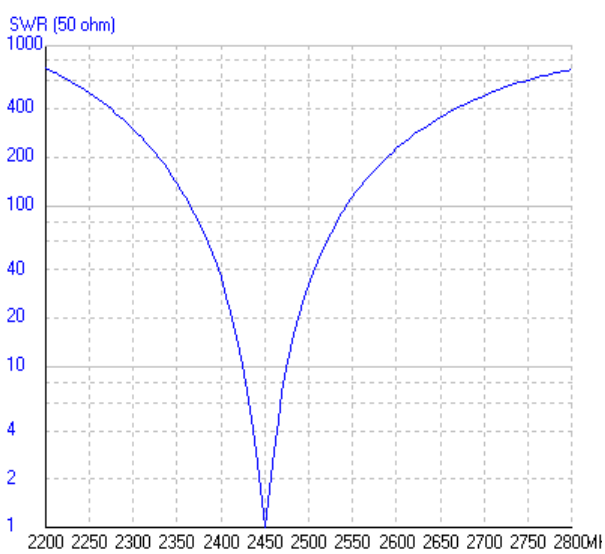


(a)

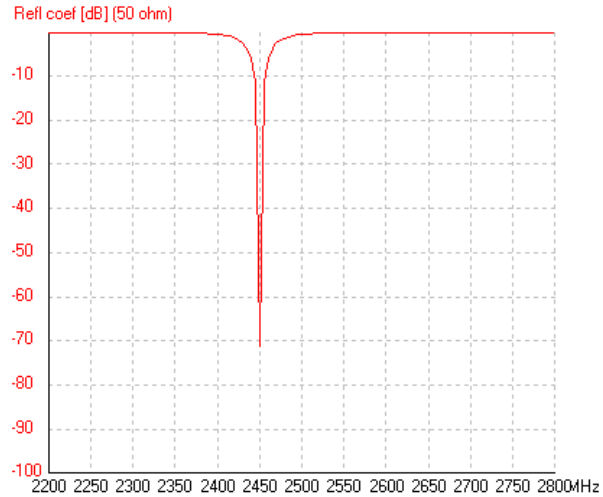


(b)

Fig. 2. (a) SWR and (b) Reflection Coefficient in ansence of Low-pass L-network



(a)



(b)

Fig. 3. (a) SWR and (b) Reflection coefficient after using low-pass L-network

The far field radiation pattern has shown in Figure 4 in which the horizontal gain & vertical gain are exactly same (22.3 dBi) whereas the totals gain of 22.26 dBi. These values are useful for RFID applications.

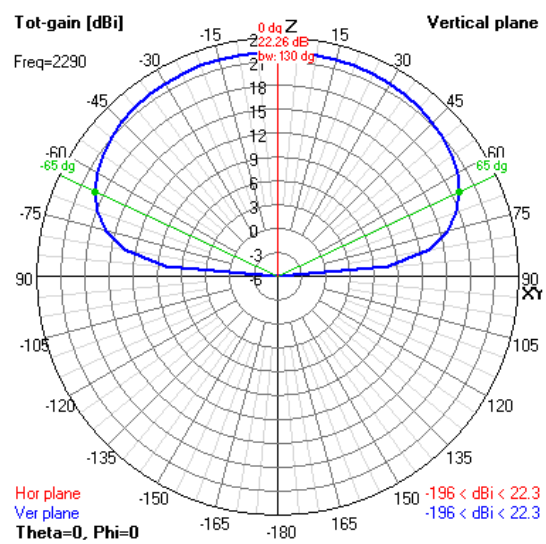
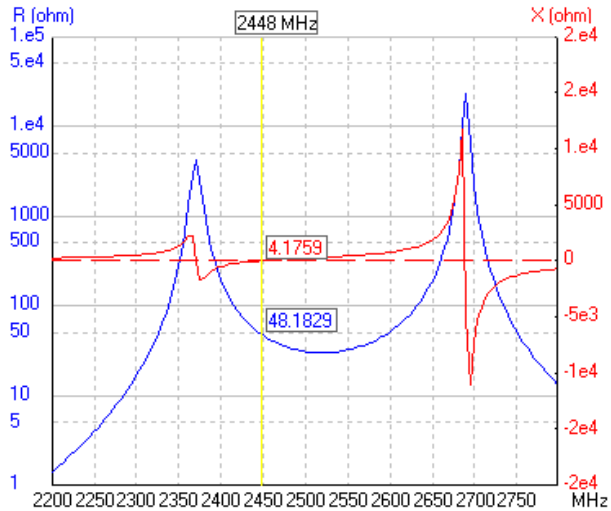
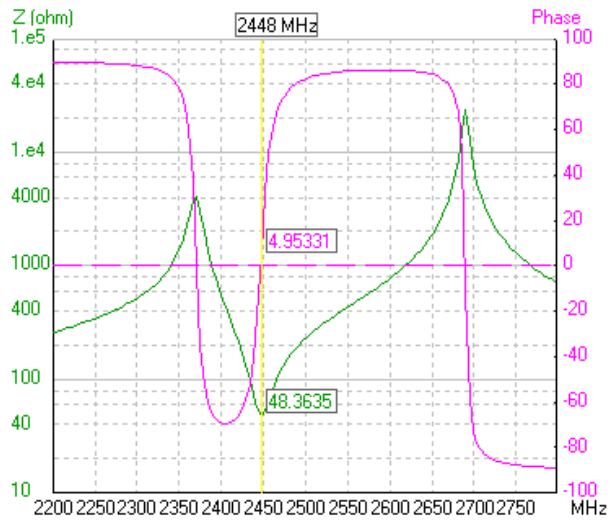


Fig. 4. Far Field Radiation Pattern

Before filtering we have got the impedance is 48.36-j4.95 ohm whereas after using the low pass L-network filter the impedance is achieved 49.99-j0.029 ohm (nearly ideal value 50 ohm) which are shown in Figure 5 & 6 respectively at frequency response closely to 2.45 GHz.

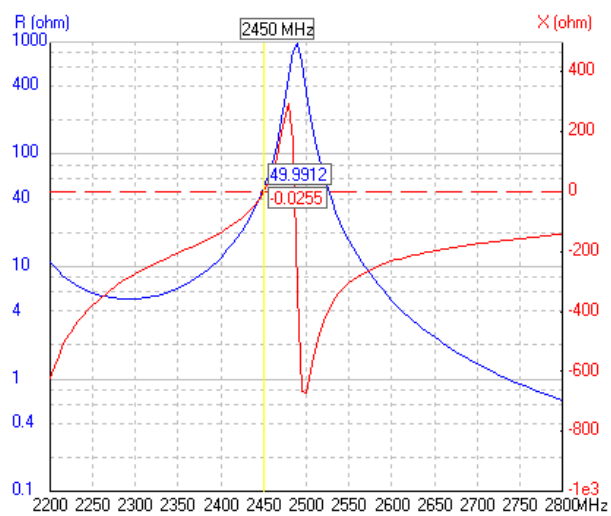


(a)

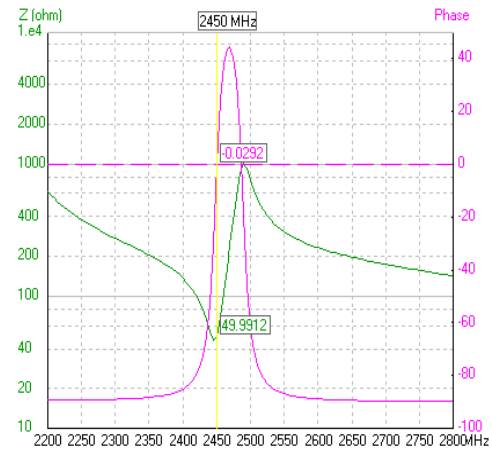


(b)

Fig. 5. (a) Resistance and reactance in ohm and (b) Magnitude of Impedance phase in degrees in absence of low-pass L-network.



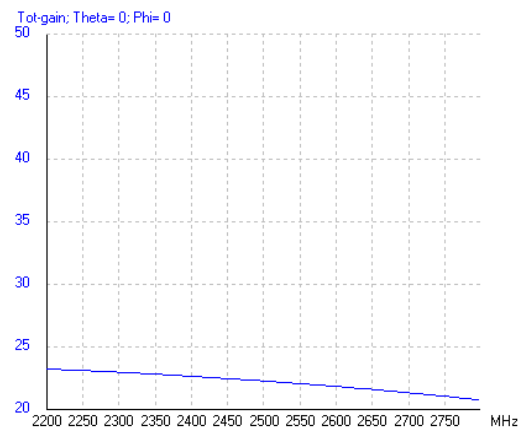
(a)



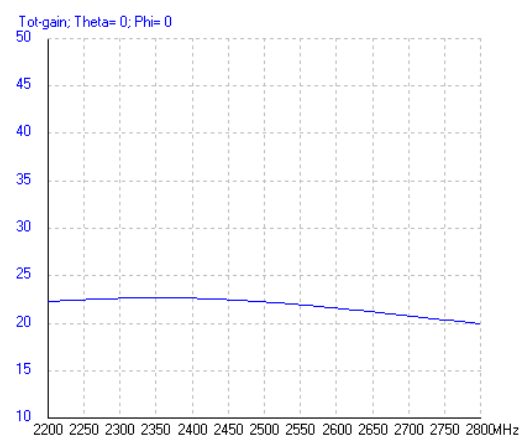
(b)

Fig. 6. (a) Resistance and reactance in ohm and (b) Magnitude of Impedance phase in degrees after using low-pass L-network.

The gain of proposed antenna has sufficiently high 22.49 dBi at response 2.45 GHz without using low pass L-network filter which is shown in Figure 6 but from Figure 7 when used low pass L-network then it lightly decrease to 22.48.



(a)



(b)

Fig. 7. (a) Antenna Gain before filtering (b) Antenna Gain after Filtering

The proposed antenna is grounded by copper sheet manufactured in the laboratory with the proposed antenna dimensions. We have found its SWR, return loss, impedance are almost near the simulated results. Here are the radiation pattern and SWR which we measured in the laboratory after being tested. The experimental radiation pattern and VSWR graph are shown in Figure 10 and Figure 11 accordingly.

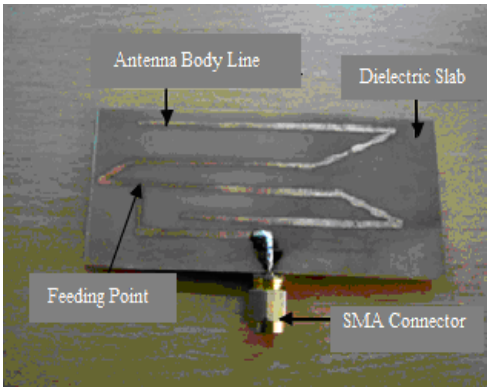
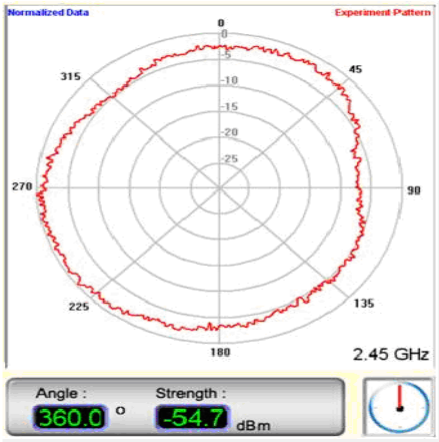
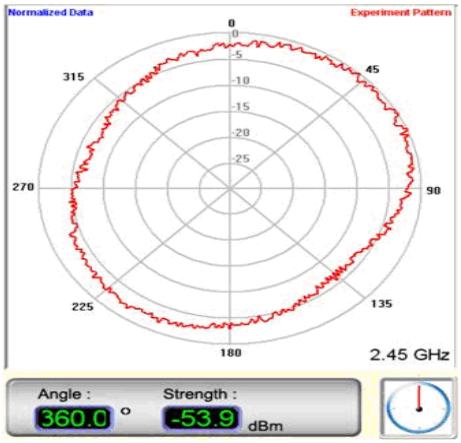


Fig. 8. Fabricated structure of W-shaped Antenna Grounded by Copper Sheet

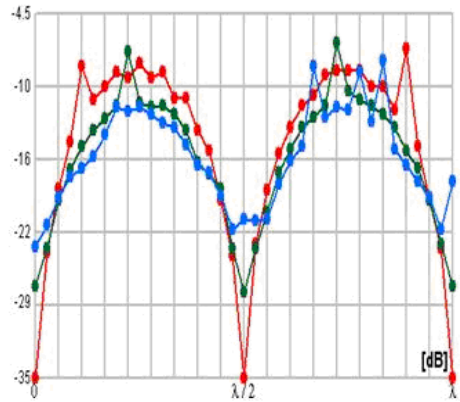


(a)

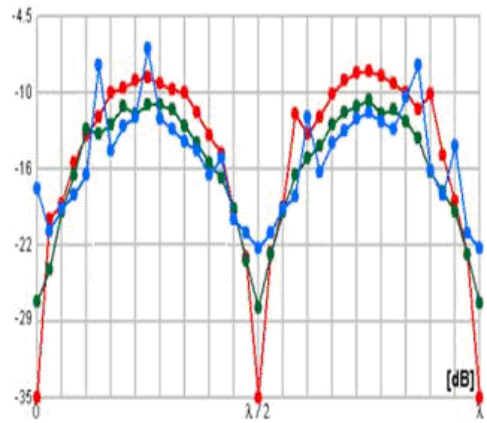


(b)

Fig. 9. Radiation Pattern in (a) Horizontal Plane and (b) Vertical Plane



(a)



(b)

Fig. 10. VSWR in (a) Vertical plane and (b) Horizontal Plane

The experimental and simulation result of the antenna are shown in Table 2 and 3 respectively. Here measured the SWR at different level of attenuations. The almost best SWR value is found at 0 dB attenuation level whereas the highest SWR value achieved at -8 dB attenuation level. But the simulation result with filtering is more decent than others and actually it is 1.00054.

TABLE II. COMPARISON OF SWR VALUES FROM MEASUREMENTS RESULTS

Attenuation	Vmax	Vmin	SWR
0 dB	-8	-35	1.05404
-4 dB	-11	-28	1.094676297
-8 dB	-12	-23	1.127648932

TABLE III. COMPARISON OF SWR VALUES FROM SIMULATION RESULTS

SWR(Simulation)	
Without Filtering	With Filtering
1.097	1.00054

IV. CONCLUSION

In this paper, we proposed a compact 2.45GHz RFID tag antenna that printed on FR4 substrate with copper cladding. We have analyzed this antenna by using Numerical Electromagnetic Code (NEC). The printed antenna structure can provide excellent characteristics and all the properties make this tag antenna design a good choice in UHF RFID applications. Recently there are worldwide many studies concerning to RFID technology for a wide range of applications. The major demands are low cost, robust, flexible, reliable, low-power consumption and durable wireless modules such as FID-enabled sensor nodes is driven by several applications, such as logistics, Aero-ID, anti-counterfeiting, supply-chain monitoring, space, healthcare, pharmaceutical, military and is regarded as one of the most important methods for realizing ubiquitous ad-hoc networks. RFID tags may be located on anything such as on humans, animals and different shopping items etc. Another thing is that there are other RFID applications in other technologies like GPS, Telephones, and Payment systems.

ACKNOWLEDGEMENTS

The authors are grateful to Mr. Numan-Al Mobin, Department of Electronics and Communication Engineering, Khulna University of Engineering and Technology for his encouragement and well support to learn the policy of simulating software NEC.

REFERENCES

[1] S.-Y. Chen, and P. Hsu, "CPW-Fed Folded-Slot Antenna for 5.8-GHz RFID Tags," Electron. Lett., vol. 40, no. 24, pp. 1516-1517, Nov. 2004.

[2] K. R. Mahmoud, "Design optimization of a BOW-TIE Antenna for 2.45 GHz RFID readers using a hybrid BSO-NM algorithm," Proc. In Electromagnetics Research, PIER 100, 105-117, 2010.

[3] S.-Y. Chen and P. Hsu, "Hybrid Loop Antenna for 2.45 GHz RFID Tags," Proc. In ISAP2005.

[4] L. H. Guo, A. P. Popov, H. Y. Li, Y. H. Wang, V. Bliznetsov, G. Q. Lo, N. Balasubramanian and D.-L.Kwong, "A Small OCA on a 1 0.5-mm2 2.45-GHz RFID Tag—Design and Integration Based on a CMOS-Compatible Manufacturing Technology," IEEE Electronic Device Letters, VOL. 27, NO. 2,pp.96-98, FEB. 2006.

[5] Li Yang, S. Serkan Basat, and Manos M. Tentzeris, "Design and Development of Novel Inductively Coupled RFID Antennas".

[6] Alberto Vargas and Lukas Vojtech, "Near Field On Chip RFID Antenna Design," ISBN 978-953-7619-72-5, pp. 324, February 2010, INTECH, Croatia.

[7] Zhiguang Fan, Shan Qiao, Jiangtao Huangfu, and Lixin Ran, "A Miniaturized 2.45GHz RFID Tag Antenna Using Planar Impedance Transformer," PIERS ONLINE, VOL. 3,pp-951-954, NO. 7, 2007.

[8] Abu Md. Numan-Al-Mobin, Md. Masud Rana, Debabrata Kumar Karmokar and Khaled Mahbub Morshed, "Compact E-Shaped Antenna for UHF RFID Transponder," International Journal of Wireless Communication and Information Systems (IJWCIS) Vol 1,pp-1-5, No 2 Oct, 2011.

[9] Randy Bancroft (Centurion Wireless Technologies) Westminster, Colorado, "Fundamental Dimension Limits of Antennas".

Heart Function Determination by Analyzing Sympathetic and Vagal Activity of ECG Signal after Having Energy Drinks

Md. Nafiur Rahman Protik[1], Fatema Khatun[2]
Dept. of Telecommunication and Electronic Engineering
Hajee Mohammad Danesh Science and Technology University, Dinajpur, Bangladesh
[1]protik69@hotmail.com, [2]fatema.tee.08@gmail.com

Muhammad Muinul Islam
Dept. of Biomedical Engineering
Khulna University of Engineering and Technology, Khulna-9203, Bangladesh
mmi.kuet@gmail.com

***Abstract*—In this work, sympathetic and vagal activity are the two component of heart rate variability (HRV) measurements of ECG signal were analyzed to determine the effect of having energy drinks on heart function by observing 12 healthy human subjects. After 6 min observation in both conditions it was found that after having energy drinks sympathetic activity was increased and vagal activity was decreased. From the previous research it is shown that heart rate is increased by slow acting sympathetic activity or decreased by fast acting parasympathetic (vagal) activity. Significant differences of sympathetic and vegal activities in both condition i.e., before and after having drinks are same i.e., 0.0045. So that to determine the effect of having energy drinks on heart function sympathetic-vagal balance was analyzed by the two component of sympathetic and vagal activity. It was observed that sympathetic-vagal balance was increased that indicates R-R interval increased as well as decreases the heart rate. Raw RR techogram also used to analyze the interval of RR. After having energy drinks a significant change also observed in sympathetic-vagal balance.**

Keywords—AcqKnowledge software; energy drinks; heart rate variability; sympathetic-vagal balance; RR tachogram

I. INTRODUCTION

Autonomic Nervous System (ANS) is the part of nervous system that non-voluntarily controls all organs and systems of the body and maintains the body under stable conditions (homeostasis). ANS consists of two subsystems: Sympathetic Nervous System (SNS) and Parasympathetic Nervous System (PNS). The sympathetic nervous system supply to the heart leaves the spinal cord at the first four thoracic vertebras, and supplies most of the muscle of the heart. Stimulation via the cardiac beta-1 receptors causes the heart rate to increase and beat more forcefully. SNS helps to prepare the body for action. PNS (vagal), on the other hand, is more active under unchallenging situations, such as rest and digestion. The vagus nerve also supplies the atria, and stimulation causes the heart rate to decrease (bradycardia). Surgical procedures can cause vagal stimulation and produce severe bradycardia. Under normal conditions the vagus nerve is the more important influence on the heart. This is especially noticeable in athletes who have slow heart rates. PNS tends to work in opposite direction of SNS, bringing the body towards a rest state. SNS activation increases heart rate and breathing rate, but PNS decreases it [1]. Constant interaction of these two systems is reflected in HRV. Analysis of variations in the instantaneous heart rate time series using the beat-to-beat RR-intervals (the RR tachogram) is known as Heart Rate Variability (HRV) analysis [2]. Therefore, HRV can be used to estimate the activity and activation level of both systems. HRV measurement reflects the balance between sympathetic and parasympathetic (vagal) input to the heart. Decreased variability suggests decreased vagal input and increased sympathetic input, which predict increased risk of arrhythmias and mortality.

HRV assists in performing frequency domain analysis of human ECG to extract standard HRV measures. HRV analysis has been shown to provide an assessment of cardiovascular disease [3]. The heart rate may be increased by slow acting sympathetic activity or decreased by fast acting parasympathetic (vagal) activity. The balance between the effects of the sympathetic and parasympathetic systems, the two opposite acting branches of the autonomic nervous system (ANS), is referred to as the sympatho-vagal balance and is believed to be reflected in the beat-to-beat changes of the cardiac cycle [4,5]. The decrease (slow acting) in sympathetic tone and increase (fast acting) in parasympathetic drive results in the restoration of sympathetic-vagal balance which reduces excessive blood pressure and improves cardiac structure and function [6].

Today energy drinks become the latest craze and a permanent fixture in our culture. An energy drink is a beverage that contains some form of legal stimulant and/or vitamins which are supposed to give consumers a short term boost in energy. Make a mental note that while the Food and Drug Administration (FDA) is allowing companies to sell and market their energy drink products, there is still very little research that has been done on them. It is suspected that the FDA allows them to be added simply because they do not pose any immediate danger to energy drink consumers [7].

The aim of the study is to determine Heart function by analyzing sympathetic and vagal activity of ECG signal after having energy drinks. Sympathetic and vagal activity was analyzed both from PSD and raw RR tachogram of recorded ECG signal to determine the Heart function after having energy drinks.

II. Materials and Methods

A. ECG Signal

The heart is divided into four chambers that are connected by heart valves. The lower two chambers of the heart are called ventricles. They are separated by a septum into the left ventricle and the right ventricle. The upper two heart chambers are called atria. Atria receive blood returning to the heart from the body and ventricles pump blood from the heart to the body. Due to the force needed to pump blood to the body, ventricles have thicker walls than do atria [8]. Fig. 1 shows the cross sectional view of heart.

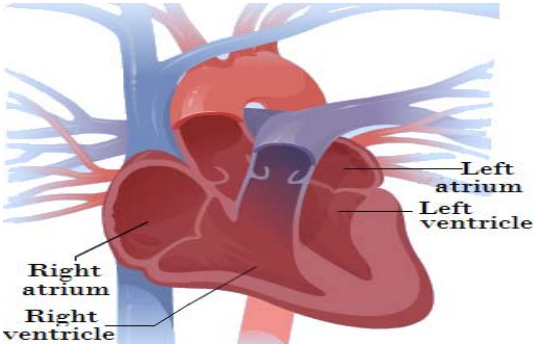


Fig. 1. Cross sectional view of heart [9]

B. Subjects

Between 24 and 32 years old Twelve healthy, male subjects were enrolled and their Mean ±SD (range) for age (23.33±2.87yrs), weight (63.67±10.61kgs), height (165.89±8.03cms) and Body Mass Index (BMI) (22.27±3.30kg/m^2) were taken where value (N=12)[a]. No medications were taken during the week prior to the study. None of the subjects were smokers and they refrained from alcohol and caffeine containing drinks at least 4 h prior to the study or had a history of cardiovascular disease or other illness. Exclusion criteria included: obesity (body mass index >30 kg/m2), under weight (body mass index <18.5 kg/m2) and history of cardiovascular disease or other illness. After being informed of the study design, they gave their written consent. The study was approved by the local Ethics Committee.

a.Values are Mean ±SD.

C. Energy Drinks

Energy drinks, unlike "sports" drinks that contain water, sugar and salts, have high levels of caffeine and taurine aimed at increasing "energy" or alertness. The “Magical” Ingredients of energy drinks have one thing in common: all of them contain a lot of sugar and/or caffeine. These could be considered the “active ingredients”.

In this experiment we have used drinks of serving size of 250ml/can which contains caffeine 62.5mg/250ml of, 27 gm sugar per 250ml can and other ingredients e.g., water, acidity regulators: citric acid and sodium citrate, carbon dioxide, taurine (0.4%), flavor, inositol (0.02%), colours: E 150d, riboflavin, enriching substances: vitamins (niacin, panthotenic acid, vitamin B6, vitamin B12).

D. Experimental Setup

The study was performed in a quiet room with the temperature kept at 22°C (21-23). The subjects were resting in the supine position throughout the whole experimental period. ECG is performed with ECG100C amplifier directly connected to biopac data acquisition unit, MP150 (Biopac, USA) through universal interface module, UIM100C (Biopac, USA).

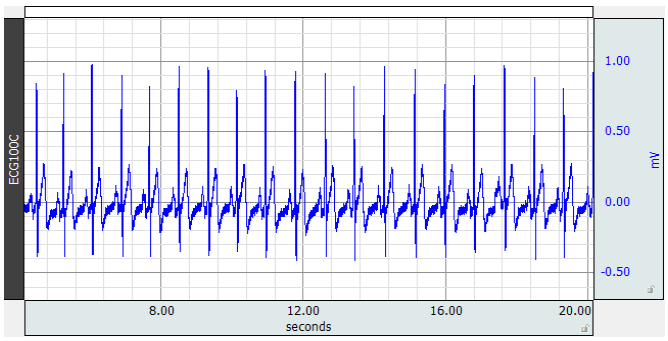
E. Testing Procedure

Food intake was restricted to a light meal 2 hour prior to the test. ECG measurements were carried out in a room in which the temperature was maintained constant at 22°C (21-23) with the subjects in a supine position. At least 20 minutes were allowed for acclimatization before the measurements were performed on the skin of middle finger tip. ECG was measured immediately before and after the taking of energy drinks for 30 minutes. Each participant had an initial visit to the experimental laboratory, for a physical examination and a medical history assessment.

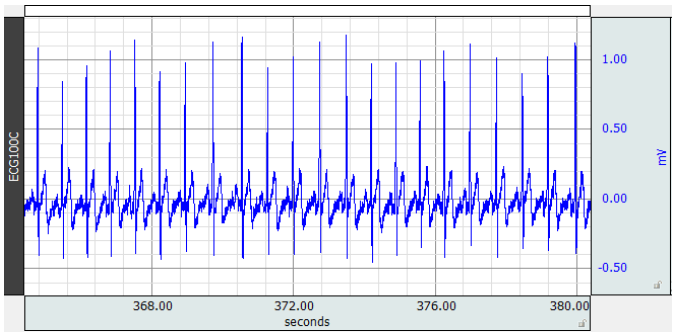
III. Result

A. Recording and Analyzing of ECG Signal

The ECG records the electrical activity of the heart, where each heart beat is displayed as a series of electrical waves characterized by peaks and valleys. Any ECG gives two kinds of information. One is the duration of the electrical wave crossing the heart which in turn decides whether the electrical activity is normal or slow or irregular and the second is the amount of electrical activity passing through the heart muscle which enables to find whether the parts of the heart are too large or overworked [9]. The pulse train of 15 sec observation of ECG wave before and after taking energy drinks is shown in Fig. 2(a) and 2(b) respectively. At normal condition, the maximum amplitudes (peak value of R wave) of ECG waves are nearer or below 1 mV. After having energy drinks, the maximum amplitudes (peak value of R wave) of ECG waves are happened around 1.1 mV i.e., greater than 1mV.



(a)

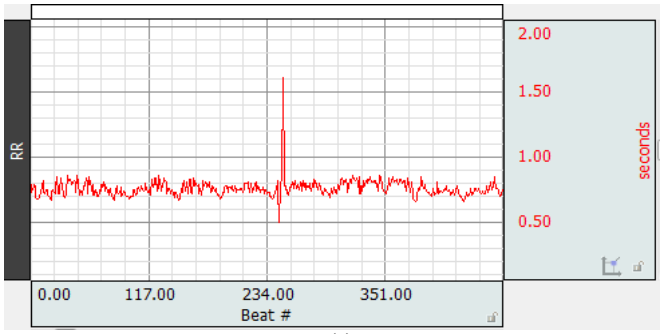


(b)

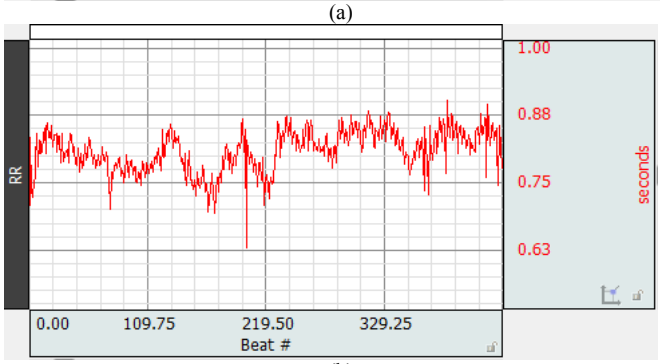
Fig. 2. ECG wave pulse traiin (a) before having energy drinks (b) after having energy drinks

B. *Analyzing Sympathetic-Vagal Balance*

The temporal difference between two peaks of 'R' wave provides the RR interval (RRI) [10], which is also known as interbeat interval, and is inversely proportional to heart rate. Spectral analysis of the RR tachogram is typically used to estimate the effect of the sympathetic and parasympathetic (vagal) modulation of the RR-intervals. HRV RR tachogram is shown at Fig. 3 before and after having energy drinks from 6 min observation.



(a)



(b)

Fig. 3. Showing raw RR techogram before and after having energy drinks with 6 min observation

From Fig. 3(a) it is shown that maximum average RR interval occur at the time of 0.759s, whereas in Fig. 3(b) average RR interval is 0.808s after having energy drinks. The average RR intervals are also shown in TABLE I. Also the activities in both cases are shown there. TABLE I shows the activity values which also mention the condition of heart activity. Biopac Acqknowledge software was used to compute the power spectrum density (PSD) of the heart period intervals. Sympathetic (low frequency) and vagal (high frequency) components of PSD were used to compute sympathetic-vagal balance.

TABLE I. AVERAGE R-R INTERVAL AND THE ACTIVITY FROM RAW RR TACHOGRAM

Event (Total Observation in Min)	MIN of Obser-vation	Before Having Energy Drinks	After Having Energy Drinks	Action
Average R-R interval (6)	6	0.759 s	0.808 s	Heart rate decreased
Sympathetic (6)	3	0.464825	0.549122	Fast acting
	3	0.287914	0.572801	Fast acting
Vagal (6)	3	0.535175	0.450878	Slow acting
	3	0.712086	0.427199	Slow acting
Sympatho-Vagal Balance (6)	3	0.868547	1.2179	Heart rate decreased
	3	0.404324	1.34083	Heart rate decreased

The two main frequency bands of interest are referred to as the Low-Frequency (LF) band (0.04 to 0.15 Hz) and the High-Frequency (HF) band (0.15 to 0.4 Hz) [11]. Sympathetic tone is believed to influence the LF component whereas both sympathetic and parasympathetic activity has an effect on the HF component. The ratio of the power contained in the LF and HF components has been used as a measure of the sympatho-vagal balance [11]. Several consequences issue from the mathematics of sympathovagal balance. First, a change of either the numerator or the denominator must change the fraction. Thus, a reduction of LF with no change of HF will reduce the fraction, and a reduction of HF with no change of LF will increase the fraction. According to published precedents [12-14], the former change would be interpreted as a shift of sympatho-vagal balance toward vagal predominance and the latter as a shift of sympatho-vagal balance toward sympathetic predominance. Second, because calculations of normalized LF and HF powers involve merely rearrangement of the same LF and HF terms, a change of normalized LF power must necessarily be associated with a change of normalized HF power. For the same reason, relations among the different calculated values can be predicted simply on the basis of mathematics. Changes in LF: HF Ratio as Markers of Changes in Sympatho-Vagal Balance. It has been a rather simple but traditional working hypothesis that conditions which increase sympathetic activity decrease vagal tone and vice versa [15]. It has recently been suggested that spectral analysis of R-R interval variability might reflect this balance [16]. In our study, the LF: HF ratio appeared to be a convenient index of such interaction. Indeed, the simultaneous decrease in vagal and increase in sympathetic mechanisms were reflected by increasing of LF: HF ratio at supine position. From TABLE I it is shown that sympathetic activity increases and vagal activities decreases and this is happen for both conditions i.e., before and after having energy drinks for every observation. Sympathetic-vagal balances are also increases as well.

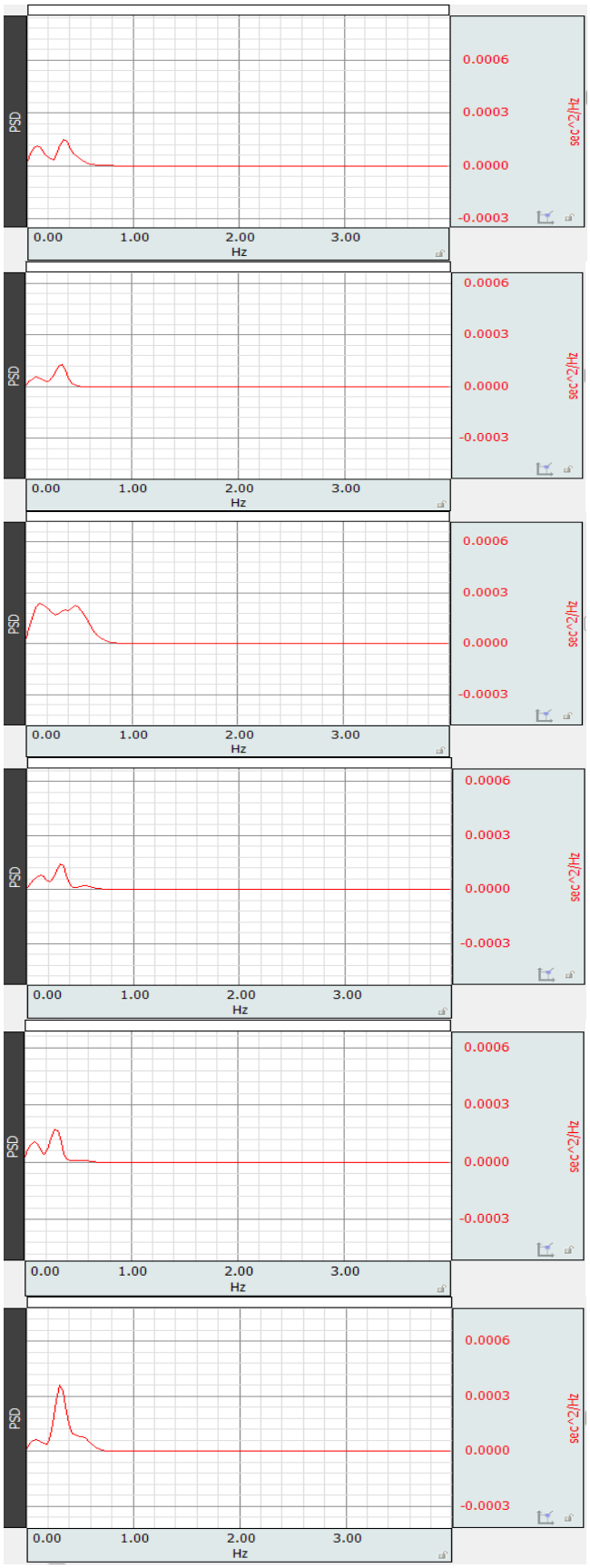


Fig. 4. HRV PSD spectrum of 1 min before having energy drink

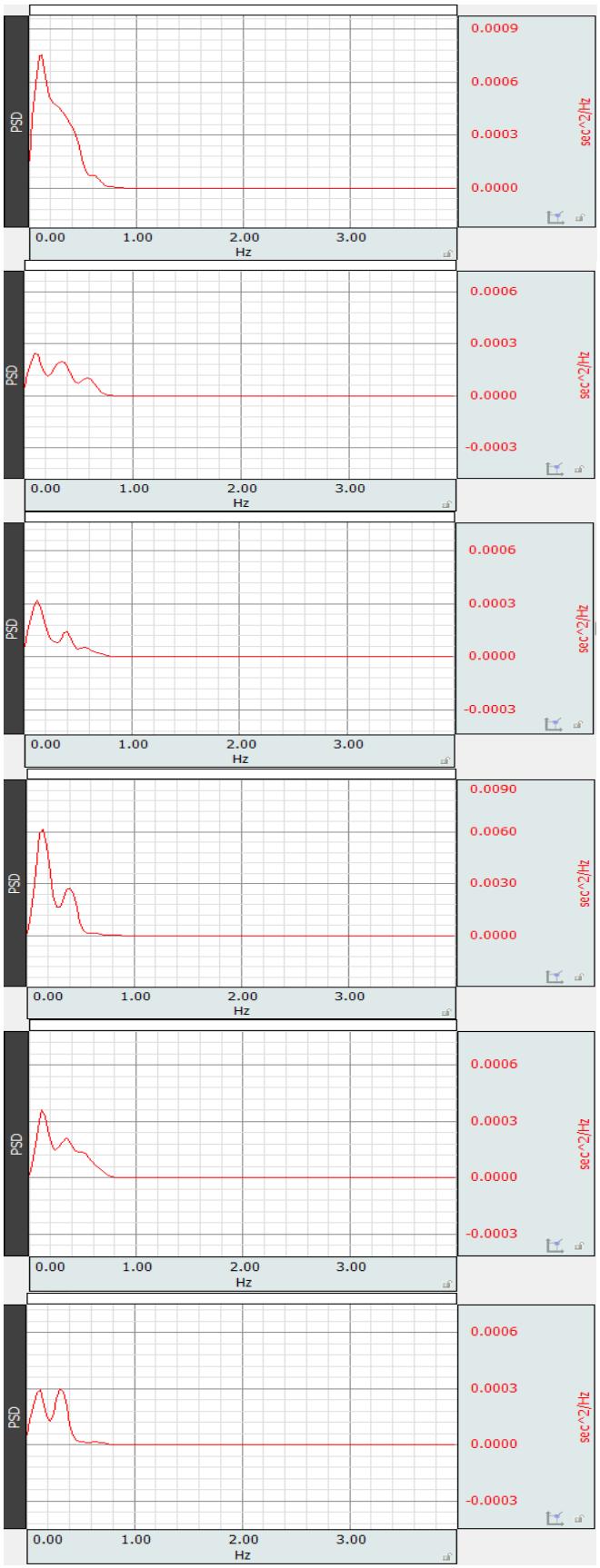


Fig. 5. HRV PSD spectrum of 1 min after having energy drink

HRV Power Spectral Density (PSD) of 6 minutes observation is shown at Fig. 4 and at Fig. 5 before and after having energy drink. Between 0 to 1 Hz at normal condition the average maximum PSD peak happened at approximately 0.0002 sec^2/Hz which shown in Fig. 5 and after having energy drinks the average maximum PSD peak happened at approximately 0.0013 sec^2/Hz which shown in Fig. 5. Displays the power spectrum density (PSD) estimation from which the PSD summations and sympathetic/vagal ratios are computed. Comparing between two conditions on each activity is shown in Fig. 6 where Fig. 6(a) stands for sympathetic activity, Fig. 6(b) for vagal activity and Fig. 6(c) for sympatho-vagal balance.

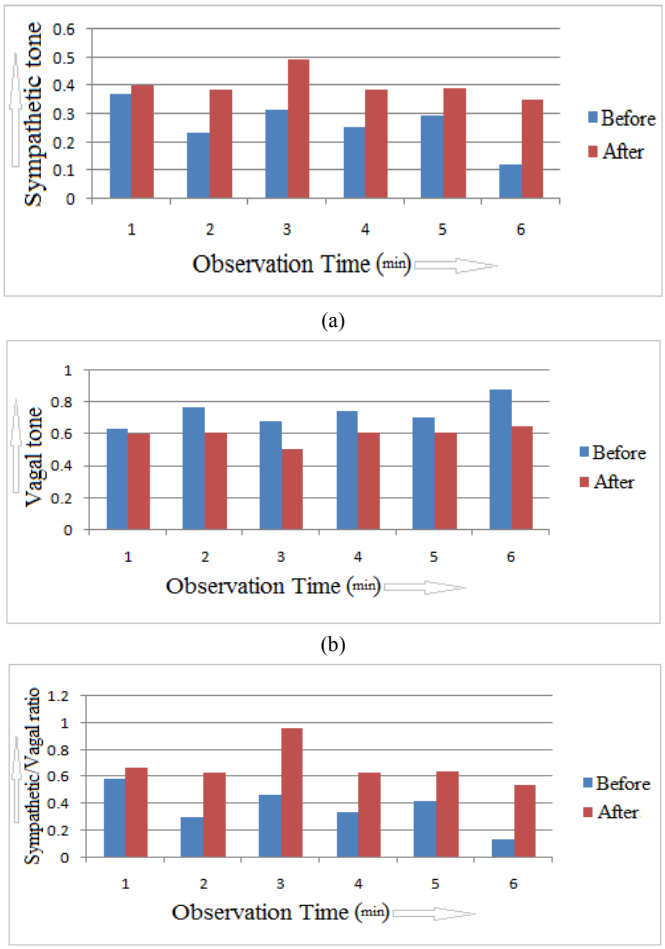


Fig. 6. Column charts for comparing (a) sympathetic activity (b) vagal activity (c) sympatho-vagal balance before and after having energy drinks

So from TABLE I it can be showed that sympathetic activity is increased at the condition of after having energy drinks which means that sympathetic activity getting slow and increasing heart rate. Also after having energy drinks vegal activity is decreased that means vegal activity getting fast and decreasing heart rate. Also it is mentioned that the decrease (slow acting) in sympathetic tone and increase(fast acting) in parasympathetic drive results in the restoration of sympathetic-vagal balance which reduces excessive blood pressure and improves cardiac structure and function and vice versa. Here the ratio also gets higher at after having energy drinks i.e., 0.375515 to 0.681150. So from this observation it can be determined clearly that heart rate decreased. Also from the increasing ratio which reflects R-R interval matter that shows greater R-R interval time and give result of slow heart rate. Because from previous research it is shown that greater R-R interval shows the slow heart rate.

C. *Statistical Analysis*

Significant differences also showed at TABLE II. For calculating P- Values we use the graph prism software. Prism summarizes the P- Value using the words in the middle column of the given data table. These P- Values are summarized in TABLE II. P- Values of before and after having energy drinks for sympathetic, vagal and Sympathetic-vagal balance are showed here. Paired t-test is taken here because it is a dependent subject. At 95% confidence level we get the p-value by two tailed paired T-test. Paired t-test is taken here because it is a dependent subject. Data is derived from study subjects who have been measured at two time points (so each individual has two measurements). This p-value shows the significant level which is very significant before and after having energy drinks.

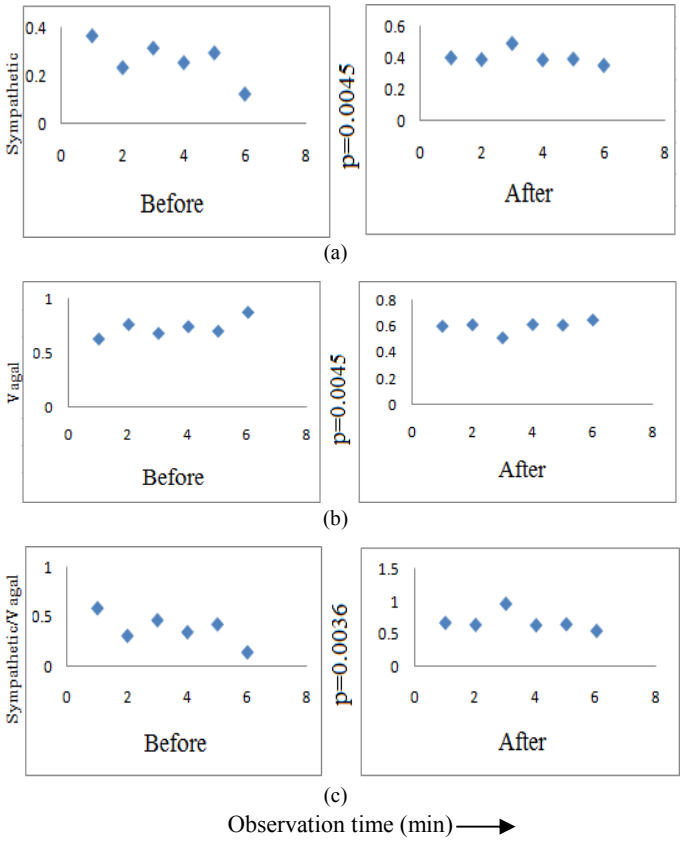


Fig. 7. Showing scatter plots of (a) sympathetic activity (b) vagal activity (c) sympatho-vagal balance with p-values before and after having energy drinks

Fig. 7(a, b, c) shows the cluster plot of three activities at both conditions- Fig. 7(a) is for sympathetic, 7(b) is for vagal and

7(c) is for sympatho-vagal balance. From Fig. 7(a) and 7(b), the measured significance differences are same i.e., p=0.0045 [see also TABLE II] and from Fig. 7(c) the significance difference is p=0.0036 [see also TABLE II].

TABLE II. SIGNIFICANCE MEASUREMENTS AND THE ACTIVITY MEASUREMENTS FROM THE PSD

(Activities)	Average 6 Min Observation			P-Values of Before v After Having Energy Drinks	Significant (Two Tailed Paired Test) P < 0.05
	Before	After	Action		
Sympathetic	0.265189	0.401917	Fast	0.0045	Yes(**)
Vagal	0.734811	0.598023	Slow	0.0045	Yes(**)
Sympatho-Vagal Balance	0.375515	0.681150	Heart rate decreased	0.0036	Yes(**)

** for very Significant

IV. DISCUSSION

In this paper, heart rate variability was analyzed to determine the effect of having energy drinks on heart function. To determine the heart function, three activities were observed: sympathetic activity, vagal activity and sympathetic-vagal balance. These activities are extracted from raw RR tachogram and from PSD. PSD and RR tachogram are applied on recorded signal in before and after having energy drinks. After observing and measuring, this study demonstrates that heart rate as well as heart function decreases. That is determined by increasing the sympathetic and sympathetic-vagal balance activity and by decreasing the vagal activity. Decreasing of heart function was also determined by the measurement of average RR interval. So the causes of heart rate decreasing are summarized below:

- Sympathetic activity increases
- Vagal activity decreases
- Sympathetic-vagal balance increases and is believed to be reflected in the beat-to-beat changes of the cardiac cycle. So beat-to-beat changes i.e., RR interval is increased.
- Average RR interval increases measured from raw tachogram and
- Significance level shows the differences for every case are very significant.

V. CONCLUSION

To analyze the effect of after having energy drinks on heart function we analyze the component and HRV of ECG signal. By analyzing them we have shown that, energy drinks has the effect on heart activity i.e., heart rate is decreased. Also it is observed that very significant differences are happened on sympathetic, vagal and sympathetic-vagal balance activity between the condition of normal and after having energy drinks.

ACKNOWLEDGMENT

Authors wish to thank all participants in this study and cordially grateful to the Department of Biomedical Engineering, Khulna University of Engineering & Technology, Khulna, where the measurements were done. The authors would also like to thanks for their most constructive remarks and suggestions.

REFERENCES

[1] Ainul Anam Shahjamal Khan, Umme Mumtahina and Nusrat Yeasmin," Heart Rate Variability Analysis Using Approximate Entropy and Detrended Fluctuation for Monitoring Heart Condition" 2nd International Conference on Informatics, Electronics & Vision (ICIEV-2013), page-89, 17-18 May, 2013

[2] De Boer RW, Karemaker JM, Strackee J: Comparing spectra of a series of point events, particularly for heart rate variability data. IEEE Trans Biomed Eng 1984;31:384-387

[3] Kitney RI, Fulton T, McDonald AH, Linkens DA: Transient interactions between blood pressure, respiration and heart rate in man. J Biomed Eng 1985;7:217-224

[4] Koizumi K, Terui N, Kollai M: Effect of cardiac vagal and sympathetic nerve activity on heart rate in rhythmic fluctuations. J Auton Nerv Syst 1985;!2:251-259

[5] Levy MN: Sympathetic-parasympafhetic interactions in the heart. Circ Res 1971 ;29:437-445

[6] Ewing DJ, Neilson JMM, Travis P: New method for assessing cardiac parasympathetic activity using 24-hour electrocardiograms. Br Heart J 1984;52:396

[7] Brian, 2004, November 23rd, "Energy Drinks: Ingredients & Dangers", Citynet Magazine. http://www.citynetmagazine.com/nightlife/energy-drinks.html

[8] C. Saritha, V. Sukanya, Y. Narasimha Murthy, "ECG Signal Analysis Using Wavelet Transforms" Bulg. J. Phys. 35 (2008) 68–77

[9] Md. Nafiur Rahman, Fatema Khatun, Md. Dulal Haque, Muhammad Muinul Islam, "Determination of the Effect of Having Energy Drinks on Tachycardia and Bradycardia by Analyzing HRV of ECG Signal" 1st International Conference on Electrical, Computer and Telecommunication Engineering (ICECTE-2012 RUET), page-72, 01-02 December, 2012

[10] Pan J and Tompkins WJ. A real –time QRS Detection algorithm.IEEE transaction on biomedical engineering 32(3):230-236,1985.

[11] Website:http://www.physionet.org/physiotools/ecgsyn/paper/node8.html #eursoccard96

[12] Pagani M, Lombardi F, Guzzetti S, Rimoldi O, Furlan R, Pizzinelli P, Sandrone G, Malfatto G, Dell'Orto S, Piccaluga E, Turiel M, Baselli G, Cerutti S, Malliani A. Power spectral analysis of heart rate and arterial pressure variabilities as a marker of sympatho-vagal interaction in man and conscious dog. Circ Res. 1986;59:178–193

[13] Brown TE, Beightol LA, Koh J, Eckberg DL. Important influence of respiration on human RR interval power spectra is largely ignored. J Appl Physiol. 1993;75:2310–2317.

[14] Malliani A, Pagani M, Lombardi F, Cerutti S. Cardiovascular neural regulation explored in the frequency domain. Circulation. 1991;84:482–492.

[15] Wesseling KH, Settels JJ, Van Der Hoeven GMA, Nijboer JA, Butijn MWT, Dorlas JC: Effects of peripheral vasoconstriction on the measurment of blood pressure in a finger. Cardiovasc Res 1985;19:139-145

[16] Pomeranz B, Macaulay RJB, Caudill MA, Kutz I, Adam D, Gordon D, Kilborn KM, Barger AC, Shannon DC, Cohen RJ, Benson H: Assessment of autonomic function in humans by heart rate spectral analysis. Am J Physiol 1985;248: H151-H153

Analyzing QRS Complex, ST Segment and QT Interval of ECG Signal to Determine the Effect of Having Energy Drinks on Hypercalcaemia

Md. Nafiur Rahman Protik[1], Fatema Khatun[2]
Dept. of Telecommunication and Electronic Engineering
Hajee Mohammad Danesh Science and Technology University, Dinajpur, Bangladesh
[1]protik69@hotmail.com, [2]fatema.tee.08@gmail.com

Muhammad Muinul Islam
Dept. of Biomedical Engineering
Khulna University of Engineering and Technology, Khulna-9203, Bangladesh
mmi.kuet@gmail.com

Abstract— **In this work, the components of ECG signal are QRS complex, QT interval and ST segment were analyzed to determine the effect of having energy drinks on hypercalcaemia by observing 12 healthy human subjects. The reasons for occurring hypercalcaemia which are analyzed here are: absence of the ST segment, wider QRS complex and shortened of QT interval. The normal interval for QRS complex is 0.04 to 0.12 seconds. After having energy drinks it is observed that there is no change happen at the interval level of QRS complex. It is cleared by the significant different which shows "ns" that means significant different does not occur at these conditions i.e., before and after having energy drinks. Also the average value of QRS complex is also smaller than 0.1 s. From this observation it is shown that one of the condition for hypercalcaema is the absence of the ST segment which is not present here i.e., ST segment is present. A little change occurs in the QT interval but it does not give any effect on the abnormality of Hypercalcaemia.**

Keywords- AcqKnowledge software; energy drinks; QRS complex; QT interval; ST segment

I. INTRODUCTION

The recording of the heart's electrical activity is known as the ECG. The deviations in the normal electrical patterns indicate various cardiac disorders. Cardiac cells, in the normal state are electrically polarized. Their inner sides are negatively charged relative to their outer sides. These cardiac cells can lose their normal negativity in a process called depolarization, which is the fundamental electrical activity of the heart. This depolarization is propagated from cell to cell, producing a wave of depolarization that can be transmitted across the entire heart. This wave of depolarization produces a flow of electric current and it can be detected by keeping the electrodes on the surface of the body. Once the depolarization is complete, the cardiac cells are able to restore their normal polarity by a process called re-polarization. This is also sensed by the electrodes [1].

Hypercalcaemia is also known as Hypercalcemia is defined as a higher-than-normal level of calcium in the blood usually more than 10.5 milligrams per deciliter of blood. It is the most common life-threatening metabolic (blood) disorder associated with cancer. Hypercalcemia is a disorder that most commonly results from malignancy or primary hyperparathyroidism. Hypercalcemia can cause a number of nonspecific symptoms, including anorexia, nausea, polydipsia, fatigue, muscle weakness, restlessness, and confusion [2].

Hypercalcemia seemed to decrease atrial activity and increase ventricular activity, as evidenced by the appearance of bradycardia, sinus arrest, and Premature Ectopic Beats. Vagal activity seemed altered by the changes in plasma calcium concentrations and involved in the production of the arrhythmia, because atropine could abolish the arrhythmia created by the infusion of the calcium solutions [3]. Hypercalcemia may produce ECG abnormalities related to altered trans-membrane potentials that affect conduction time. QT interval shortening is common secondary to absence of the ST segment. The QoT, QaT, and QeT intervals which are measured from the beginning of the QRS complex to the origin (O), apex (A), and the end (E) of the T wave, respectively are very short. At very high levels, the QRS interval may lengthen, T waves may flatten or invert and a variable degree of heart block may develop. The electrocardiographic abnormalities in combined hypercalcaemia are: absence of the ST segment, prolonged T wave, a shortened QT interval and prominent U waves [4].

Today energy drinks become the latest craze and a permanent fixture in our culture. An energy drink is a beverage that contains some form of legal stimulant and/or vitamins which are supposed to give consumers a short term boost in energy. Make a mental note that while the Food and Drug Administration (FDA) is allowing companies to sell and market their energy drink products, there is still very little research that has been done on them. It is suspected that the FDA allows them to be added simply because they do not pose any immediate danger to energy drink consumers [5].

The aim of this study is to determine the effect of having energy drinks on Hypercalcaema by analyzing ECG signal's QRS intervals. For this here we analyze the ECG signal components and specially analyze the QRS complex i.e., interval between QRS onset and QRS end, ST segment, QT interval.

II. Materials and Methods

A. ECG Signal

The heart is divided into four chambers that are connected by heart valves. The lower two chambers of the heart are called ventricles. They are separated by a septum into the left ventricle and the right ventricle. The upper two heart chambers are called atria. Atria receive blood returning to the heart from the body and ventricles pump blood from the heart to the body. Due to the force needed to pump blood to the body, ventricles have thicker walls than do atria [6]. Fig. 1 shows the cross sectional view of heart.

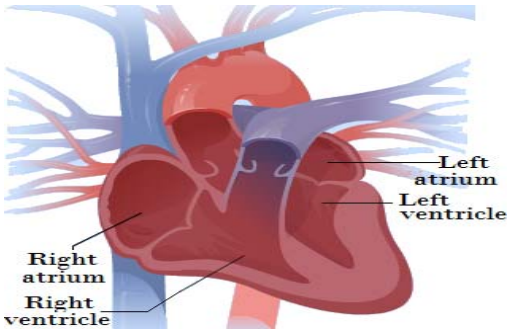


Fig. 1. Cross sectional view of heart [7]

Fig. 2 shows a normal ECG waves. ECG signal is characterized by five peaks and valleys labeled by the letters P, Q, R, S, T. In some cases we also use another peak called U. The P-wave represents the activation of the upper chambers of the heart, the atria, while the QRS complex and T-wave represent the excitation of the ventricles- the lower chamber of the heart.

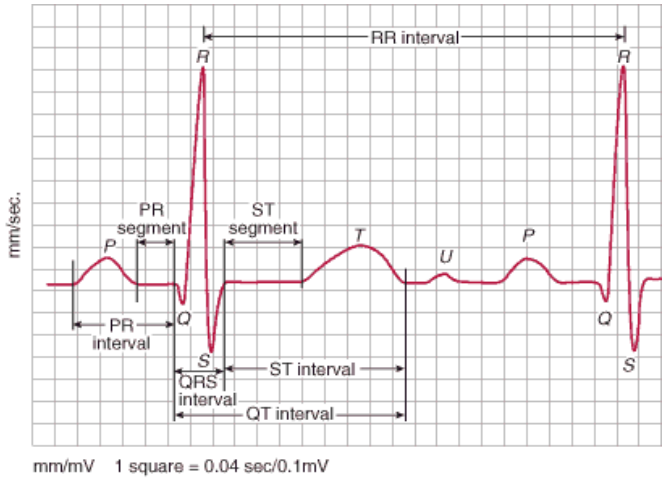


Fig. 2. Normal ECG waves [8].

P wave = activation (depolarization) of atria.
QRS complex = depolarization of ventricles, consisting of the Q, R, and S waves.
QT interval = time interval between onset of ventricular depolarization and end of ventricular repolarization.
R-R interval = time interval between 2 QRS complexes.
T wave = ventricular repolarization.
ST segment plus T wave (ST-T) = ventricular repolarization.
Normally, the frequency range of an ECG signal is of 0.05–100 Hz and its dynamic range of 1–10 mV. This ECG is used clinically in diagnosing various abnormalities and conditions associated with the heart. In the normal sinus rhythm (normal state of the heart) the P-R interval is in the range of 0.12 to 0.2 seconds. The QRS interval is from 0.04 to 0.12 seconds. The Q-T interval is around 0.42 seconds and the normal rate of the heart is from 60 to 100 beats per minute. So, from the recorded shape of the ECG, we can say whether the heart activity is normal or abnormal. Any change in the above said values indicates the abnormality of the heart. Characteristics feature for the abnormality of hypercalcaema is shown in TABLE I. Among all the waves, QRS complex is the most significant wave of an ECG signal [9].

TABLE I. Characteristic Feature for the Abnormality of Hypercalcaemia

Name of abnormality	Characteristic feature (occur when)
Hypercalcaemia	QRS interval > 0.1 s
	Shortening of QT interval
	Absence of the ST segment

B. QRS Complex Morphology

Normal QRS width is 70-100 ms (duration of 110 ms is sometimes observed in healthy subjects). The QRS width is useful in determining the origin of each QRS complex (e.g., sinus, atrial, junctional or ventricular). Narrow complexes (QRS < 100 ms) are supraventricular in origin. Broad complexes (QRS > 100 ms) may be either ventricular in origin, or may be due to aberrant conduction of supraventricular complexes (e.g., due to bundle branch block, hyperkalaemia or sodium-channel blockade).
Broad Complexes:

- A QRS duration > 100 ms is abnormal
- A QRS duration > 120 ms is required for the diagnosis of bundle branch block or ventricular rhythm

Broad complexes may be ventricular in origin or due to aberrant conduction secondary to Bundle branch block, Hyperkalaemia, Poisoning with sodium-channel blocking agents (e.g., tricyclic antidepressants), Pre-excitation (i.e., Wolff-Parkinson-White syndrome), Ventricular pacing, Hypothermia, Intermittent aberrancy (e.g., rate-related aberrancy)[10].

C. Subjects

Between 24 and 32 years old Twelve healthy, male subjects were enrolled and their Mean ±SD (range) for age (23.33±2.87yrs), weight (63.67±10.61kgs), height (165.89±8.03cms) and Body Mass Index (BMI) (22.27±3.30kg/m^2) were taken where value (N=12)[a]. No medications were taken during the week prior to the study. None of the subjects were smokers and they refrained from alcohol and caffeine containing drinks at least 4 h prior to the

study or had a history of cardiovascular disease or other illness. Exclusion criteria included: obesity (body mass index >30 kg/m2), under weight (body mass index <18.5 kg/m2) and history of cardiovascular disease or other illness. After being informed of the study design, they gave their written consent. The study was approved by the local Ethics Committee.

a.Values are Mean ±SD.

D. *Energy Drinks*

Energy drinks, unlike "sports" drinks that contain water, sugar and salts, have high levels of caffeine and taurine aimed at increasing "energy" or alertness. The "Magical" Ingredients of energy drinks have one thing in common: all of them contain a lot of sugar and/or caffeine. These could be considered the "active ingredients".

In this experiment we have used drinks of serving size of 250ml/can which contains caffeine 62.5mg/250ml of, 27 gm sugar per 250ml can and other ingredients e.g., water, acidity regulators: citric acid and sodium citrate, carbon dioxide, taurine (0.4%), flavor, inositol (0.02%), colours: E 150d, riboflavin, enriching substances: vitamins (niacin, panthotenic acid, vitamin B6, vitamin B12).

E. *Experimental Setup*

The study was performed in a quiet room with the temperature kept at 22°C (21-23). The subjects were resting in the supine position throughout the whole experimental period. ECG is performed with ECG100C amplifier directly connected to biopac data acquisition unit, MP150 (Biopac, USA) through universal interface module, UIM100C (Biopac, USA).

F. *Testing Procedure*

Food intake was restricted to a light meal 2 hour prior to the test. ECG measurements were carried out in a room in which the temperature was maintained constant at 22°C (21-23) with the subjects in a supine position. At least 20 minutes were allowed for acclimatization before the measurements were performed on the skin of middle finger tip. ECG was measured immediately before and after the taking of energy drinks for 30 minutes. Each participant had an initial visit to the experimental laboratory, for a physical examination and a medical history assessment.

III. RESULT

A. *Recording and Analyzing of ECG Signal*

The ECG records the electrical activity of the heart, where each heart beat is displayed as a series of electrical waves characterized by peaks and valleys. Any ECG gives two kinds of information. One is the duration of the electrical wave crossing the heart which in turn decides whether the electrical activity is normal or slow or irregular and the second is the amount of electrical activity passing through the heart muscle which enables to find whether the parts of the heart are too large or overworked. The pulse train of 15 sec observation of ECG wave before and after taking energy drinks is shown in Fig. 3(a) and 3(b) respectively. At normal condition, the maximum amplitudes (peak value of R wave) of ECG waves are nearer or below 1 mV.

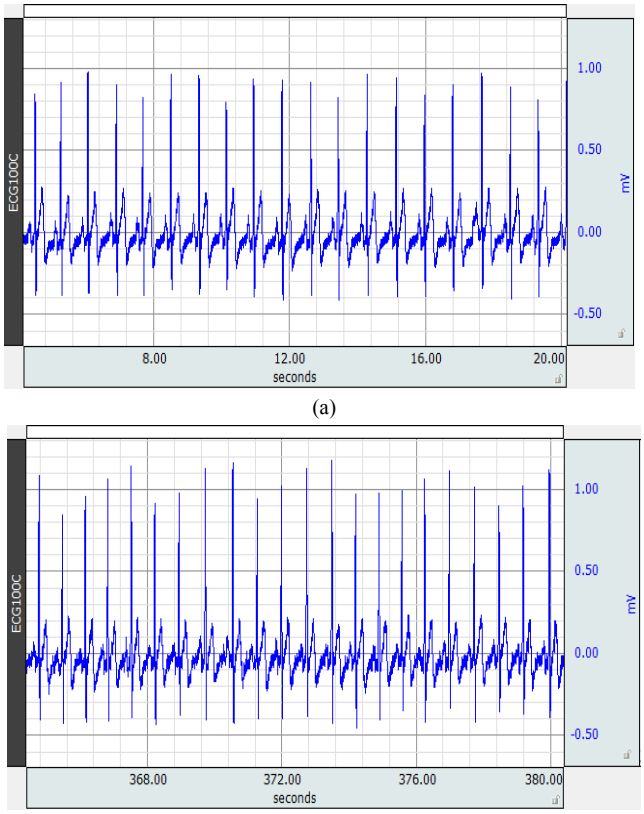
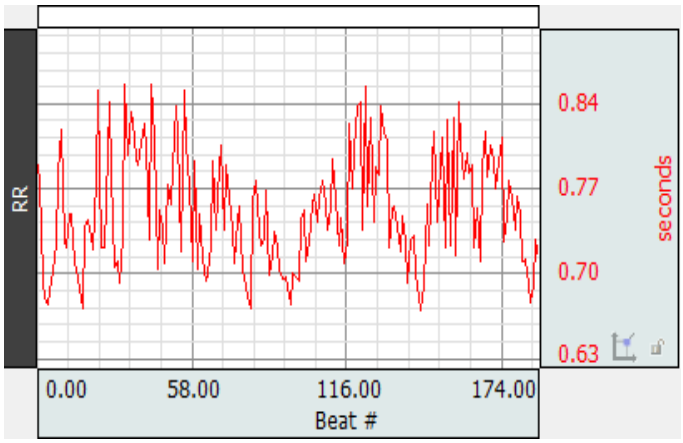


Fig. 3. ECG wave pulse traiin (a) before taking energy drinks (b) after taking energy drinks

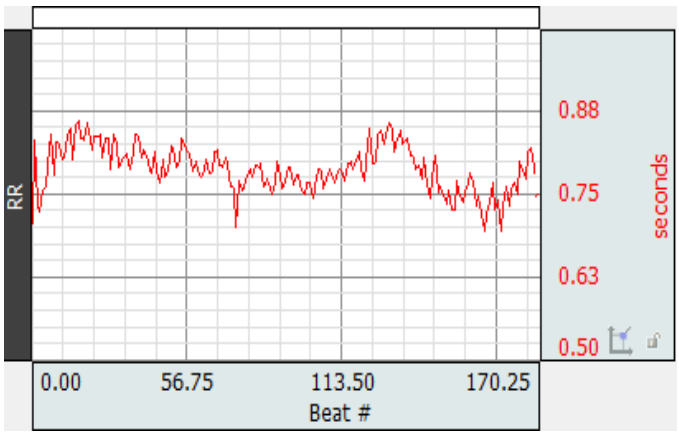
After having energy drinks, the maximum amplitudes (peak value of R wave) of ECG waves are happened around 1.1 mV i.e., greater than 1mV. Not only R peak but also there is a significant increase in Q peak & S peak in ECG signal after having energy drinks. From TABLE II we observed that, the maximum amplitude of R wave in ECG signal increases about 0.14 mV after having energy drinks with respect to normal condition.

TABLE II. PEAK AMPLITUDE OF R(QRS PEAK) WAVE

Peak amplitude *of R wave* (mV)	
Before Having Energy Drink	**After Having Energy Drink**
0.96	**1.08**
0.94	**1.11**
0.95	**1.10**
0.93	**1.16**
0.96	**1.10**
0.94	**1.17**
0.92	**1.09**



(a)



(b)

Fig. 4. RAW RR tachogram (a) before having energy drinks(b) after having energy drinks

The temporal difference between two peaks of 'R' wave provides the RR interval (RRI) [11], which is also known as interbeat interval, and is inversely proportional to heart rate. Spectral analysis of the RR tachogram is typically used to estimate the effect of the sympathetic and parasympathetic (vagal) modulation of the RR-intervals. Fig. 4(a,b) shows the HRV RR tachogram those are shown at before and after having energy drinks from 6 min observation. From Fig. 4(a) it is shown that maximum average RR interval occur at the time of 0.759s, whereas in Fig. 4(b) average RR interval is 0.808s after having energy drinks.

B. *Analyzing QRS Complex of ECG Signal*

The performance of ECG analyzing system depends mainly on the accurate and reliable detection of the QRS complex, as well as T- and P waves. The duration of the QRS complex is another useful calculation. The detection of the QRS complex is the most important task in automatic ECG signal analysis. Once the QRS complex has been identified a more detailed examination of ECG signal including the heart rate, the ST segment *etc.* can be performed [12].

The duration of the QRS complex is another useful calculation. Once the QRS complex has been identified a more detailed examination of ECG signal including the heart rate can be performed [13].

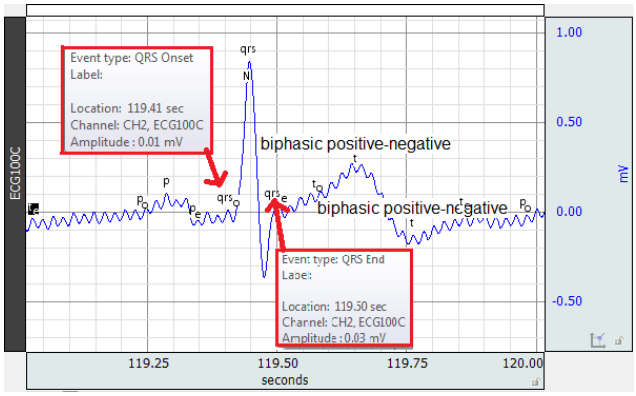


Fig. 5. Showing QRS complex onset and end point from recorded ECG signal

Fig. 5 shows the QRS complex onset and end point from recorded ECG signal. The QRS complex represents ventricular depolarization. The Q wave is the initial downward deflection; normal Q waves last < 0.05 sec in all leads except $V_{1\text{-}3}$, in which any Q wave is considered abnormal, indicating past or current infarction. The R wave is the first upward deflection; criteria for normal height or size are not absolute, but taller R waves may be caused by ventricular hypertrophy. A 2nd upward deflection in a QRS complex is designated R′. The S wave is the 2nd downward deflection if there is a Q wave and the first downward deflection if not. The QRS complex may be R alone, QS (no R), QR (no S), RS (no Q), or RSR′, depending on the ECG lead, vector, and presence of heart disorders. Normally, the QRS interval is 0.07 to 0.10 sec. An interval of 0.10 to 0.11 sec is considered incomplete bundle branch block or a nonspecific intraventricular conduction delay, depending on QRS morphology; ≥ 0.12 sec is considered complete bundle branch block or an intraventricular conduction delay. The average QRS complex measured from the recorded ECG signal in both conditions are showed in TABLE III. Before having energy drinks the average QRS complex interval value is 0.099605s and after having energy drinks it goes to 0.096184s.

TABLE III. AVERAGE QRS COMPLEX BEFORE AND AFTER HAVING ENERGY DRINKS

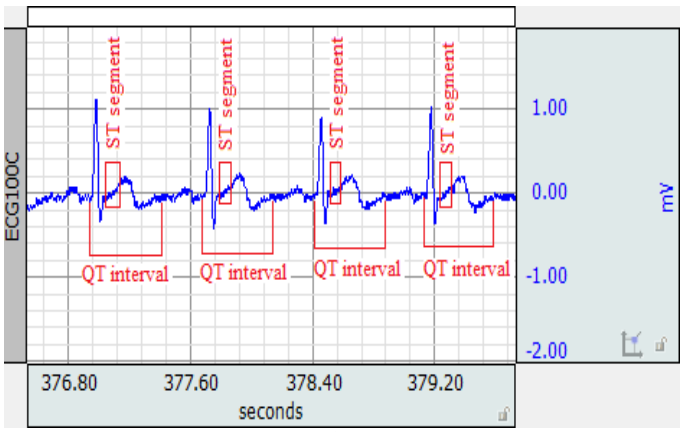
Condition	Average QRS Complex
Before	0.099605 s
After	0.096184 s

C. *Analyzing QT Interval and ST Segment of ECG Signal*

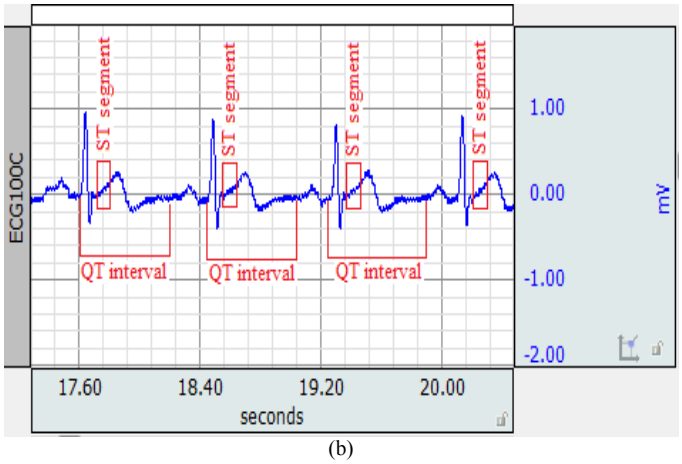
The QT interval is the time between onset of ventricular depolarization and end of ventricular repolarization. The QT-interval starts at the beginning of the QRS complex and finishes at the end of the T-wave. It represents the time taken for the ventricles to depolarise and then repolarise. The main

ECG abnormality seen with hypercalcaemia is shortening of the QT interval [14].

The ST-segment starts at the end of the S-wave & finishes at the start of the T-wave. It represents ventricular repolarisation. It should be level with the PR-segment and the T-P segment in healthy individuals. If it is elevated it suggests the individual is suffering a myocardial infarction. If it is depressed it suggests the presence of ischemic myocardial tissue in the ventricles.



(a)



(b)

Fig. 6. QT interval and ST segment are showing at (a) before and (b) after having energy drinks

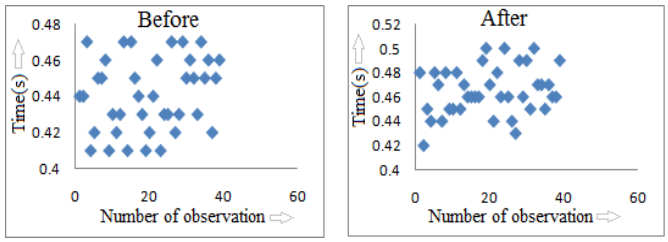
Fig. 6 shows the ST segment and QT interval region at both conditions. It also shows the presence of ST segment in every ECG signal in both conditions. So from Fig. 6 it can be say that one of the condition for hypercalcaema is absent of the ST segment is not present. From TABLE IV it is showed that QT interval goes 0.44s to 0.46s that means it is not shortened.

TABLE IV. AVERAGE QT INTERVAL AT BEFORE AND AFTER HAVING ENERGY DRINKS

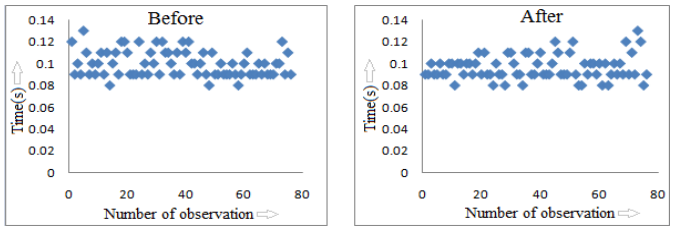
Condition	Average QT interval
Before	0.440769 s
After	0.464872 s

D. *Statistical Analysis*

Significant difference between two conditions i.e., before and after having energy drinks is showed at TABLE V. For calculating P- Values we use the graph prism software. Prism summarizes the P- Value using the words in the middle column of the given data table. Paired t-test is taken here because it is a dependent subject. At 95% confidence level we get the p-value for QRS complex is 0.0684, which is measured by two tailed paired T-test. Paired t-test is taken here because it is a dependent subject. Data is derived from study subjects who have been measured at two time points (so each individual has two measurements). This p-value shows the significant level which is very significant before and after having energy drinks. This summary of the test is shown at TABLE V. Fig. 7(a, b) shows the cluster plot of two intervals one is for QT interval and other for QRS complex interval. From Fig. 7(a) the maximum observation occurs between 0.41s- 0.47s. The maximum observation occurs between 0.42s-0.50s. So the average QT intervals calculated in before having energy drinks is 0.44s and after having energy drinks is 0.46s. Also from Fig. 7(b) QRS complex interval occurs between 0.08-0.13s which is same for both conditions i.e., before and after having energy drinks.



(a)



(b)

Fig. 7. Scatter plots of (a) QT interval (b) QRS interval before and after having energy drinks

TABLE V. SIGNIFICANCE MEASUREMENTS OF QRS COMPLEX AT BEFORE AND AFTER HAVING ENERGY DRINKS

Analyzed	Before V After Having Energy Drinks
P-value	0.0684
P value summary	Ns
Are means signif. different? (P < 0.05)	No
One- or two-tailed P value?	Two-tailed
T- Test	Paired t-test
Number of pairs	76

IV. Discussion

In this paper, among various components of ECG signal three components are analyzed to observe and determine the effect of having energy drinks on hypercalcaema. These three components are QRS complex, QT interval and ST segment. ECG signal are recorded from the subjects which are used to analyze QRS complex, QT interval and ST segment. When the interval of QRS complex goes greater than 0.1s then it shows the effect as well as abnormality named as hypercalcaema. After having energy drinks it is shown that the average time interval of QRS complex is under the interval time that is 0.1s. ST segment observation also shows the presence its existence in recorded ECG signal. QT interval has the significant difference in both conditions i.e., before and after having energy drinks. But these differences not mention any abnormality because it's time interval is belong to normal condition. Also QT interval shortening is common secondary to absence of the ST segment. So it can be said that no shortening of QT interval as well as no absence of ST segment is occurred. From the RR tachogram it is also shown that RR interval is increased so that heart beat rate is decreased after having energy drinks.

V. Conclusion

To analyze the effect of after having energy drinks on hypercalcaema three analyses are done on ECG signal. To make such decision QRS complex, QT interval and ST segment are analyzed here. By analyzing them we have shown that, energy drinks has no effect on hypercalcaema because our analysis show the results which are normal. The conditions for occurring hypercalcaema are not fulfilled after having energy drinks.

Acknowledgment

Authors wish to thank all participants in this study and cordially grateful to the Department of Biomedical Engineering, Khulna University of Engineering & Technology, Khulna, where the measurements were done. The authors would also like to thanks for their most constructive remarks and suggestions.

References

[1] M G. Tsipouras, D I Fotiadis, D Sideris , "Arrhythmia Classification using the RR-interval Duration Signal",IEEE Transactions on Computers in Cardiology, Vol. 29, 485-488, 2002.

[2] Rayner HC, Hasking DJ. Hyperparathyroidism associated with severe hypercalcaemia and myocardial calcification despite minimal bone disease. Br Med J (Clin Res Ed). 1986; 293:1277-1278

[3] Littledike ET, Glazier D, Cook HM.Electrocardiographic changes after induced hypercalcemia and hypocalcemia in cattle: reversal of the induced arrhythmia with atropine. Am J Vet Res. 1976; 37:383-388

[4] Chia BL, Thai AC. Electrocardiographic abnormalities in combined hypercalcaemia and hypokalaemia--case report. Ann Acad Med Singapore. 1998; 27:567-569

[5] Brian, 2004, November 23rd, "Energy Drinks: Ingredients & Dangers", Citynet Magazine. http://www.citynetmagazine.com/nightlife/energy-drinks.html

[6] De Boer RW, Karemaker JM, Strackee J: Comparing spectra of a series of point events, particularly for heart rate variability data. IEEE Trans Biomed Eng 1984;31:384-387

[7] Md. Nafiur Rahman, Fatema Khatun, Md. Dulal Haque, Muhammad Muinul Islam, "Determination of the Effect of Having Energy Drinks on Tachycardia and Bradycardia by Analyzing HRV of ECG Signal" 1st International Conference on Electrical, Computer and Telecommunication Engineering (ICECTE-2012 RUET), page-72, 01-02 December, 2012

[8] Website:http://www.merckmanuals.com/professional/cardiovascular_disorders/cardiovascular_tests_and_procedures/electrocardiography_ecg.html#v931664

[9] C. Saritha, V. Sukanya, Y. Narasimha Murthy, "ECG Signal Analysis Using Wavelet Transforms" Bulg. J. Phys. 35 (2008) 68–77

[10] Website:http://lifeinthefastlane.com/ecg-library/basics/qrs-complex-morphology/

[11] Pan J and Tompkins WJ. A real –time QRS Detection algorithm.IEEE transaction on biomedical engineering 32(3):230-236,1985.

[12] *Biopac Blood flow Monitor*, Biopac Systems, Inc. ISO 9001:2000, http://www.biopac.com/Manuals/laser-doppler-flow.pdf

[13] Mancia G. Autonomic modulation of the cardiovascular system during sleep. N Engl J Med 1993;328(5):347–9.

[14] Website:http://lifeinthefastlane.com/ecg-library/basics/qt_interval/

An Improved MaxProp Based on Neighborhood Contact History for Delay Tolerant Networks

Md. Nazmus Sadat
Department of Computer Science and Engineering
Bangladesh University of Engineering and Technology
Dhaka, Bangladesh
nazmus.cse@gmail.com

Muhammad Tasnim Mohiuddin
Department of Computer Science and Engineering
Bangladesh University of Engineering and Technology
Dhaka, Bangladesh
tasnimcse08@gmail.com

***Abstract*— Delay Tolerant Networks (DTNs) are intermittently connected mobile networks, in which a fully connected path from source to destination does not exist. Therefore in these networks, message delivery relies on opportunistic routing where nodes use store-carry-and-forward paradigm to route the messages. However, effective forwarding based on a limited knowledge of contact behavior of nodes is challenging. In some routing algorithms, (such as Prophet [1], MaxProp [2]) the messages are forwarded to nodes which has higher chance to meet the destination. The MaxProp was designed for vehicle based DTN. It has some drawbacks. For instance, during normalization step of updating delivery probability, it drastically decrements the delivery probability for the nodes corresponding to older encounters. On the other hand, it increments one node's delivery probability, corresponding to the latest encounter to a great extent. This procedure of updating delivery probability can suffer from non-optimal routing decisions. In this paper we propose a new technique for updating delivery likelihood based on neighborhood contact history. For the simulation, we have used Opportunistic Network Environment (ONE) Simulator. The simulation shows that our proposed technique performs much better than original MaxProp.**

Keywords*— *Delay-tolerant network; Routing; The ONE Simulator

I. INTRODUCTION

Intermittently connected mobile networks are mobile wireless networks where most of the time there does not exist a complete path from a source to a destination or such a path is highly unstable and may change or break soon after it has been discovered (or even while being discovered). This situation arises when the network is quite sparse, in which case it can be viewed as a set of disconnected, time varying cluster of nodes. Intermittently connected mobile networks belong to the general category of Delay Tolerant Networks (DTN).

Several approaches have been adopted to achieve reliable communication in such challenged networks [7], [8], [9], [11], [13], [14]. The researchers have focused on various issues like reducing the delivery delay or increasing the delivery ratio. Optimizing resource usage, providing scalability etc. are also issues explored by various algorithms.

An important DTN routing schemes is one where node mobility is not under control of the routing algorithm. Epidemic routing [3] is the most basic of these methods that replicates messages whenever a node meets another node which has not received the message so far. Prophet [1] estimates the delivery predictability to destinations using the history of encounters and the transitive property of meeting with nodes. There exist schemes where the number of times a message can be replicated is pre-specified. Example includes Spray and Wait [4], Spray and Focus [10] that limits the total number of copies created initially.

The NECTAR protocol [5] uses contact history to create a neighborhood index. The scheduler of NECTAR uses the neighborhood index (contact history) to select the message to be forwarded.

The MaxProp [2] is a protocol designed for vehicle based DTNs. Such a DTN is characterized by large storage capacity and energy source but short contact duration. Hence it discusses prioritization of packets to be forwarded and dropped. These priorities are decided by path likelihood. Thus it becomes a controlled flooding.

II. BACKGROUND

MaxProp estimates the delivery likelihood as follows:

Let the set of the nodes in the network be s. Each node, $i \epsilon s$,

keeps track of a probability of meeting peer $j \epsilon s$.The probability f_j^i estimates the likelihood that the identity of the node connects to next will be j. For all nodes, is initially set to 1/(|s|-1). When node j is encountered, the value of f_j^i is incremented by 1 and then all values are re-normalized. Each time two peers met, they exchange these values.

For example, for a DTN with four other nodes, a peer has values for f_1^i=f_2^i= f_3^i=f_4^i=0.25.Upon encountering node 3, the peer sets f_3^i=1.25 and re-normalizes these values so that they sum to 1: f_1^i=f_2^i=f_4^i=0.125 and f_3^i=0.625.

Now consider the following scenario:

This peer then encounters node1 and sets: f_1^i=1.125.Then after re-normalization f_1^i=0.5625, f_2^i=f_4^i=0.0625 and f_3^i=0.3125.

At this time, if another peer meets with the first one, it will see that f_3^i is only 0.3125.But the first peer just met node 3 and it is

very likely that they will meet again soon. So, we can conclude that, this estimation technique is not stable at all.

At the core of MaxProp is ranked list of the peer's packets based on a cost assigned to each destination .The cost is an estimate of delivery likelihood .With other node's values in hand, a local node calculates a cost (i,i+1,……..,d), for each path possible to the destination d, up to n hops long.

III. IMPROVING MAXPROP

In real stochastic scenarios, a node movement does not occur in a totally random way. During movement there is a greater likelihood that nodes meet again some of their past neighbors. The neighbors of a node X may move toward a certain destination establish contact with other nodes and may spread the information that there was a known route to node X. Nodes that learned this route may move toward other destinations and may spread the route to node X to a new set of neighbors, and so on.

We estimate the delivery likelihood based on the contact history in order to create neighborhood index and then determine the optimal route. For this purpose we use the concept of the NECTAR protocol [5]. The NECTAR protocol uses the occurrence of an opportunistic contact to calculate a neighborhood index and spread messages in a controlled manner .In real stochastic scenario nodes move in a way that there is some likelihood that these nodes can meet again their past neighbors. Consequently sending messages in a controlled way to a neighborhood of a destination increases the delivery likelihood.

The neighborhood index is calculated in such a way that nodes that are frequent neighbors present a high neighborhood index. When the contact between node i and j occurs, the neighborhood index to each other is assigned to 1, and while nodes i and j are within communication range, the neighborhood index and the contact counter are increased linearly. Then nodes i and j update the neighborhood index for destinations that are not within the communication range.

The neighborhood index formula is shown in the following equation:

$$N_{(i',d)} = \frac{Contact_{(j,d)}}{(Hops_{(j,d)}+1) \times (TS - ts_update + 1)^{\sigma}}$$

In addition, if node i has already a route to node d and node j has a better neighborhood index to node d, $N_{(i,d)}$ will be updated in a weighted fashion.

If node d is unknown, then

$$N_{(i,d)} = N_{(i',d)}$$

Otherwise,

$$N_{(i,d)} = \frac{(N_{(i,d)} \times \omega) + N_{(i',d)}}{\omega + 1}$$

TABLE I. NECTAR PARAMETERS

$Contact_{(i,j)}$	Define the amount of time slots that i and j remained in contact
$Hops_{(i,j)}$	Express the number of hops required for i to reach j
TS	Current Time Stamp
$ts_update_{(i,j)}$	Time Stamp of the last route update from i to j
α	Define the maximum value of TTL
σ	Define the aging constant
ω	Weight applied to a known neighborhood index
ɣ	Define maximum buffer occupancy for receiving messages during the Epidemic Level phase
MinEpidemicLevel	During this phase, NECTAR acts as the Epidemic Protocol
MaxEpidemicLevel	During this phase, NECTAR depends on the value of the parameter to act or not as the Epidemic Protocol
$N_{(i,j)}$	Neighborhood Index from i to destination j

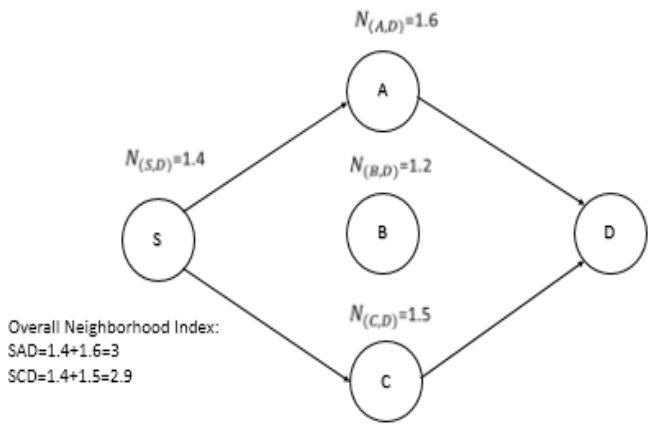


Fig. 1. Overall neighborhood index calculation.

In MaxProp the cost is an estimate of delivery likelihood. But here we replace cost by Overall Neighborhood Index which is the sum of the neighborhood indexes along the route from source to destination.

Figure 1 shows the procedure of calculating overall neighborhood indexes. Here, node S forwards message only to node A and C, because both $N_{(A,D)}$ and $N_{(C,D)}$ are greater than $N_{(S,D)}$. But $N_{(B,D)}$ is smaller than $N_{(S,D)}$.So, S does not forward packets to B.

Again, between SAD and SCD, SAD has higher Overall Neighborhood Index. So, Improved MaxProp chooses SAD to forward message.

IV. SIMULATION SETUP

Simulations play an important role in analyzing the behavior of DTN routing protocols. With sparsely distributed nodes, DTN simulations abstract from the details of wireless link characteristics and simply assume that two nodes can communicate when they are in range of one another. This allows focusing on the evaluation of the DTN routing protocols.

DTN routing and application protocol's performance is highly dependent on the underlying mobility and node characteristics. Evaluating DTN protocols across many scenarios requires suitable simulation tools. We used the Opportunistic Network Environment (ONE) simulator which is specifically designed for evaluating DTN routing algorithms.

At its core, ONE is an agent-based discrete event simulation engine. At each simulation step the engine updates a number of modules that implement the main simulation functions. The main functions of the ONE simulator are the modeling of node environment, inter-node contacts using various interfaces, routing, message handling and application interactions.

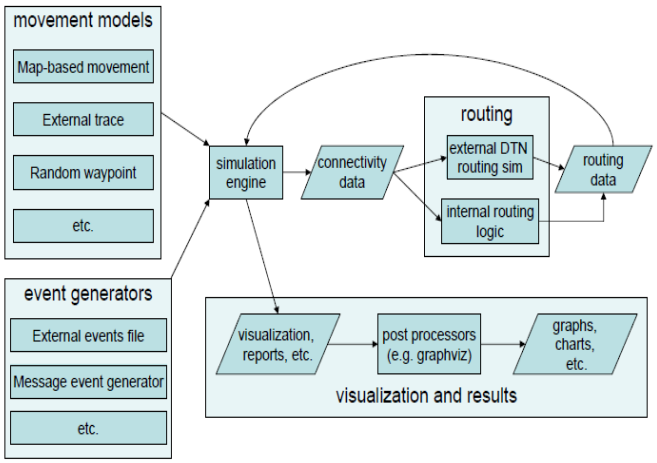


Fig. 2. Overview of ONE simulator environment [6].

The message routing capability is implemented similarly to the movement capability: the simulator includes a framework for defining the algorithms and rules used in routing and comes with ready implementations of MaxProp and other well known DTN routing protocols. To evaluate our proposed improved MaxProp in the simulator we created a new routing module.

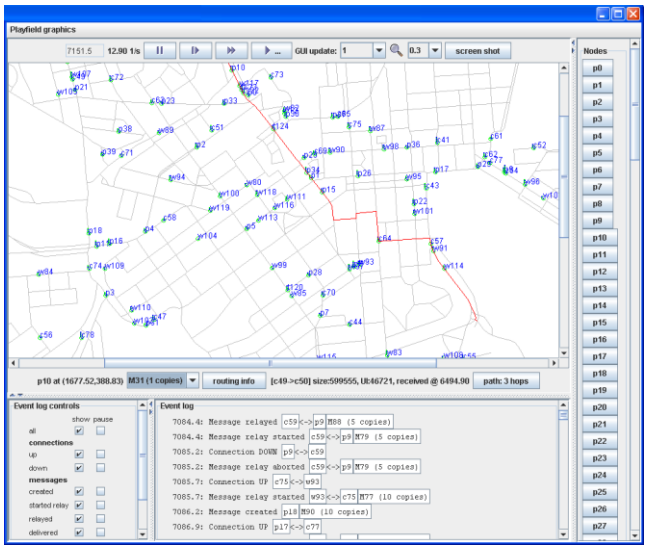


Fig. 3. Screenshot of the ONE simulator's GUI [12].

TABLE II. SIMULATION PARAMETERS

Parameter	Value
Total Simulation Time	12 Hours
World Size	4500×3400 m
Movement Model	ShortestPathMapBasedMovement
Routing Protocol	MaxProp, Improved MaxProp
Node Buffer Size	5M
No of Nodes	126
Interface Transmit Speed	2 Mbps
Interface Transmit Range	10 m
msg TTL	60,120,180,240,300,360
Node Movement Speed	Min=0.5 m/s Max=1.5 m/s
Message Creation Rate	One Message Per 25-35 sec
Message Size	500 to 1 MB

V. SIMULATION RESULTS AND EVALUATION

We have investigated the performance of MaxProp and improved MaxProp by varying message TTL. The performance is analyzed on three metrics: Delivery ratio, Overhead ratio and Average latency.

In the simulation environment, we have focused on comparing the performance with regard to the metrics mentioned above .The simulation results presented here are obtained by running the simulations as per the parameters defined in Table 2.

A. Delivery Ratio

It is the fraction of generated message that are correctly delivered to the final destination within a given time period. It is defined as: $\frac{Number\ of\ packets\ delivered}{Number\ of\ packets\ created}$

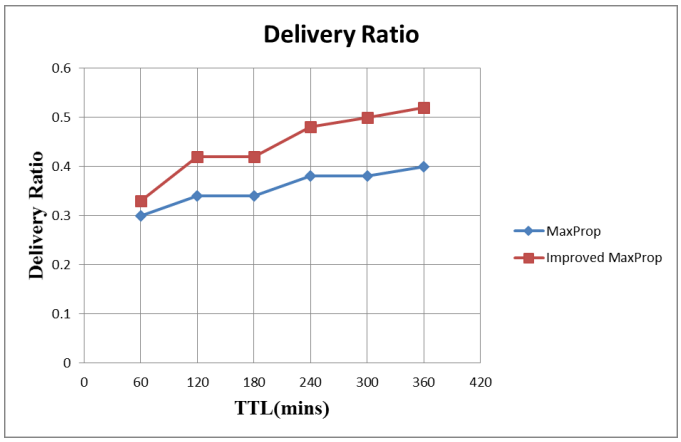


Fig. 4. Performance of Improved MaxProp in terms of Delivery Ratio.

Figure 4 compares our proposed approach's performance to original MaxProp. From this, it is evident that improved MaxProp outperforms original MaxProp in terms of delivery ratio.

B. Overhead Ratio

This metric is used to estimate the extra number of data packets needed by the routing protocol for actual delivery of the data packets. It is defined as:

$$\frac{Number\ of\ packets\ relayed - Number\ of\ packets\ delivered}{Number\ of\ packets\ delivered}$$

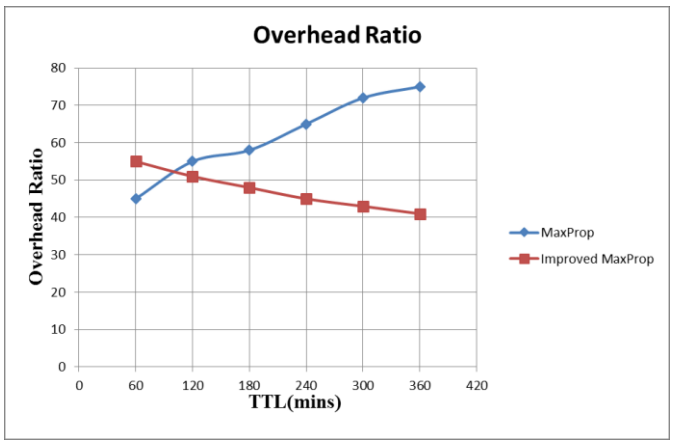


Fig. 5. Performance of Improved MaxProp in terms of Overhead Ratio.

Figure 5 compares our proposed approach's performance to original MaxProp. From this, it is evident that improved MaxProp outperforms original MaxProp in terms of overhead ratio.

C. Average Latency

It is the measure of average time between messages is generated and when it is received by the destination.

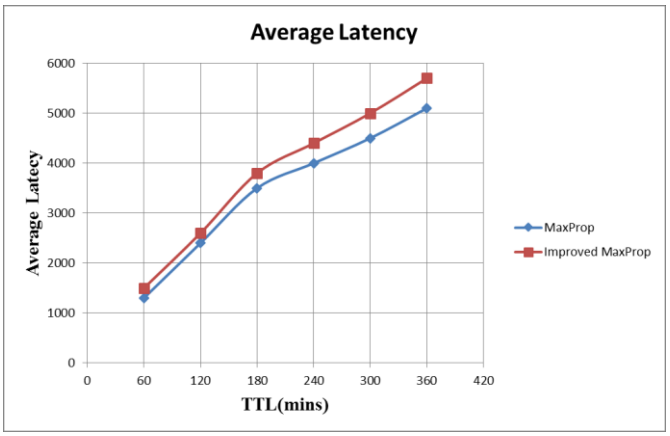


Fig. 6. Performance of Improved MaxProp in terms of Average Latency.

Figure 6 compares our proposed approach's performance to original MaxProp. Here we can see that improved MaxProp outperforms original MaxProp in terms of Average Latency too.

VI. Conclusion

In this paper we present a novel approach to estimate delivery likelihood for MaxProp routing protocol. We exploit the underlying recurrent movement pattern here.

As a future work, we intend to find a technique to make good predictions on future encounters. We will focus on introducing a few more metrics that all independently capture the network behavior from different perspectives and then merge them together.

References

[1] O. S. Andres Lindgren, Avri Doria, and O. Schelen, "Probabilistic routing in intermittently connected networks," ACM SIGMOBILE Mobile Computing and Communications Review, vol. 7, pp. 19–20, Jul. 2003.

[2] J. Burgess, B. Gallagher, D. Jensen, and B. N. Levine, "MaxProp: Routing for vehicle based disruption-tolerant networks," in Proc. IEEE International Conference on Computer Communications (INFOCOM '06), Barcelona, Spain, Apr. 2006, pp. 1–11.

[3] A. Vahdat and D. Becker, "Epidemic routing for partially connected ad hoc networks," Department of Computer Science, Duke University, Tech. Rep. CS-2000-06, April 2000.

[4] T. Spyropoulos, K. Psounis, and C. Raghavendra, "Spray and wait: an efficient routing scheme for intermittently connected mobile networks," in Proc. of ACM SIGCOMM workshop on DTNs (WDTN), Philadelphia, USA, Aug. 2005, pp. 252–259.

[5] E. C. R. de Oliveira and C´ elio V. N. de Albuquerque, "NECTAR: A DTN Routing Protocol Based on Neighbourhood Contact History," in Proc. ACM Symposium on Applied Computing (SAC'09), Hawaii, USA, Mar. 2009, pp. 40-46.

[6] Ari Keränen, Jörg Ott and Teemu Kärkkäinen, "The ONE Simulator for DTN Protocol Evaluation," in Proc. International Conference on

Simulation Tools and Techniques (SIMUTools'09), Rome, Italy, Mar. 2009.

[7] C. Liu and J. Wu, "Scalable routing in delay tolerant networks," in Proc. ACM international symposium on Mobile ad hoc networking and computing (MobiHoc'07), Montreal, Quebec, Canada, Sep. 2007, pp 51–60.

[8] Q. Li and D. Rus, "Sending messages to mobile users in disconnected ad-hoc wireless networks," in Proc. the 6th annual international conference on Mobile computing and networking (MobiCom'00), Boston, MA, Aug. 2000, pp. 44–55.

[9] A. Balasubramanian, B. Levine, and A. Venkataramani, "DTN routing as a resource allocation problem," in Proc. the 2007 conference on Applications, technologies, architectures, and protocols for computer communications (SIGCOMM '07), Kyoto, Japan, Aug. 2007, pp. 373–384.

[10] T. Spyropoulos, K. Psounis and C. Raghavendra, "Spray and focus: Efficient mobility assisted routing for heterogeneous and correlated mobility," in Proc. IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOMW '07), White Plains, NY, USA, Mar. 2007, pp. 79–85.

[11] V. Erramilli, M. Crovella, A. Chaintreau, and C. Diot, "Deligation forwarding," in Proc. ACM international symposium on Mobile ad hoc networking and computing (MobiHoc '08), Hong Kong SAR, China, May 2008, pp. 251-260.

[12] Opportunistic Network Environment (ONE) simulator, "http://www.netlab.tkk.fi/tutkimus/dtn/theone."

[13] H. Dubois-Ferriere, M. Grossglauser and M. Vetterli, "Age matters: Efficient route discovery in mobile ad hoc networks using encounter ages," in Proc. ACM international symposium on Mobile ad hoc networking & computing (MobiHoc '03), Annapolis, MD, USA, Jun. 2003, pp. 257-266.

[14] W. Zhao, M. Ammar, and E. Zegura, "Multicast routing in delay tolerant networks: Semantic models and routing algorithms," in Proc. ACM SIGCOMM workshop on Delay-tolerant networking (WDTN '05), Philadelphia, PA, USA, Aug. 2005, pp. 268-275.

The mCard approach for Bangladesh

A Smart Phone based Credit/Debit/ATM Card

Mohammad Salah Uddin,
Dipartimento di Informatica e Sistemistica
Sapienza Universit`a di Roma,
Rome, Italy
uddin@dis.uniroma1.it

Nakul Chandra Das
Dept. of Computer Science and
Engineering, East West University,
Dhaka, Bangladesh
arpanraj@live.com

Aoyan Barua
Dept. of Computer Science and
Engineering, East West University,
Dhaka, Bangladesh
aoyan_b@yahoo.com

***Abstract*— Mobile payments are a natural evolution of e-payment schemes that will facilitate mobile commerce. A mobile payment or m-payment may be defined, as any payment where a mobile device is used to initiate, authorize and confirm an exchange of financial value in return for goods and services. Mobile devices may include mobile phones, PDAs, wireless tablets and any other device that connect to mobile telecommunication network and make it possible for payments.**

In Bangladesh, several private banks (Dutch-Bangla Ltd, Islami Bank Ltd, Trust Bank Ltd and Others) introduced mobile banking. The BRAC Bank-initiated mobile banking service, "bKash", is at present the country's leading service-provider in mobile banking. This paper present a new m-payment system called "mCard" replacing existing Credit or Debit or ATM card and includes some feature of mobile banking (except money transfer to remote area).

***Index Terms*— Mobile Payment, bkash, mCard, NPS, POS, ATM, Credit/Debit Card, Android, Apps, Smart Phone.**

I. Introduction

Ten billion people are expected to own mobile phones in the globe by 2015. There are currently 99.87 million mobile subscribers in Bangladesh [1]. In a few years more than 200 million people are expected to have mobile phones in Bangladesh.

Mobile commerce is a natural successor to electronic commerce. The capability to pay electronically coupled with a website is the engine behind electronic commerce. Electronic commerce has been facilitated by automatic teller machines (ATMs) and shared banking networks, debit and credit card systems, electronic money, and electronic bill payment systems. Mobile payments are a natural evolution e-payment schemes that will facilitate mobile commerce. A mobile payment or m-payment may be defined, as any payment where a mobile device is used to initiate, authorize and confirm an exchange of financial value in return for goods and services [2]. Mobile devices may include mobile phones, PDAs, wireless tablets and any other device that connect to mobile telecommunication network and make it possible for payments to be made [3]. Music, video on demand, location based services identifiable through mobile handheld devices – procurement of travel, hospitality, entertainment and other uses are possible when mobile payments become feasible and ubiquitous. Mobile payments can be a complement to cash, cheques, credit cards and debit cards. It can also be used for payment of bills (especially utilities and insurance premiums) with access to account-based payment instruments such as electronic funds transfer, Internet banking payments, direct debit and electronic bill payments.

Several mobile payment companies and initiatives in Europe have failed and many have been discontinued [4]. In Europe and North America with few exceptions such as Austria, Spain and Scandinavian countries the development of mobile payments have not been successful. However, mobile payment services in Asia have been fairly successful especially in South Korea, Japan and other Asian countries (e.g., Mobile Suica, Edy, Moneta, Octopus, GCash). NTT DoCoMo has 20 million subscribers and 1.5 million of them have activated credit card functionality in Japan [5]. The main difference between successful implementations of mobile payment services in the Asia Pacific region and failure in Europe and North America is primarily attributed to the "payment nature" of the consumers that are country-specific.

Figure 1: The Architecture of bkash

II. EXISTING TTECHNOLOGY

A. bkash

Bangladesh's first complete mobile financial service provider, bKash Limited, a BRAC Bank subsidiary, launched its mobile banking operation in July 2011. bKash is designed to provide financial services via mobile phones to both the unbanked and the banked people of Bangladesh. The overall bKash value proposition is simple: a safe, convenient place to store money; a safe, easy way to make payments and money transfers. The bKash mobile wallet is a VISA technology platform which is fully encrypted to ensure most secure transactions, will be the customer account into which money can be deposited and out of which money can be withdrawn or used for various services. Customers will be able to receive electronic money into their bKash accounts through salary, loan, domestic remittance, and other disbursements and eventually will cash out the electronic money from bkash authorized agents/ATM [6].

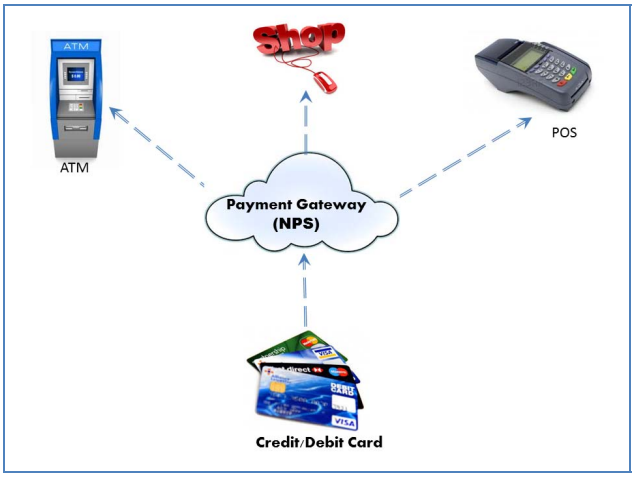


Figure 2: National Payment Switch(NPS)

B. National Payment Switch

Bangladesh bank launched National Payment Switch (shows in figure 2) in 2012, a common platform for the country's commercial banks for electronic payments, which is going to revolutionize E-commerce in Bangladesh. Once all banks in Bangladesh join the NPS (National Payment Service), a customer using a credit or debit card of any bank will be able to draw cash from any ATM and POS (point of sales) in the country, and transaction costs will significantly go down as the transactions will be routed through the NPS instead of Visa, Master or Amex card networks abroad. Moreover, purchase and sale can be done with cards through the internet and web portals within the country [7]. NPS will facilitate easy access to funds, an increase in tax revenue, more flow of money in the economy, and financial inclusion. NPS is a common electronic payment platform to help banks and their clients make transactions in an easy and cost effective way. Individual payment platforms introduced by different banks will be abolished once all banks join the common platform.

III. THE PROPOSED SMARTPHONE BASED CREDIT/DEBIT/ATM CARD SYSTEM

Android is the most popular mobile operating system for smart phone/Tabs. Now a days most of the people in Bangladesh use smart phones/Tabs. Users of smartphone/tab are increasing day by day.

Figure 3: The Proposed mCard System

A. System Overview

"mCard" is a Android apps that runs on Android based smart phone / tabs. It connects Smart phone/Tabs with National Payment System (NPS). The "mCard" replace the traditional plastic based credit/debit/ATM card. So, instead of using cards (credit/debit) we are now use (mCard) mobile apps as a virtual card. Figure 3 shows the proposed model of mCard.

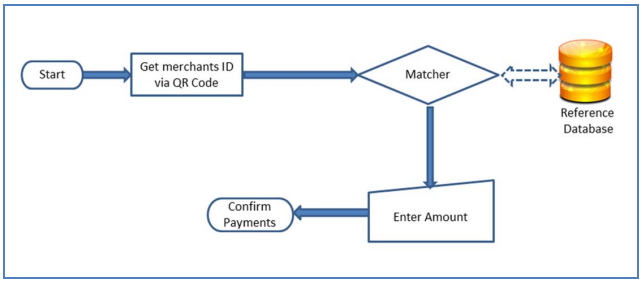


Figure 4: Flows of operation of mCard

B. Flow of Operation

A registration is mandatory before using Smart Phone as a virtual card (mCard). The mCard application is available to the bank's website. Since it replaces the traditional plastic card (Debit/Credit/ATM), the documents that are require for mCard is same as Credit/Debit card according to the bank's rules and regulation. mCard system uses IMEI [11] number (the unique identifier of mobile phone) as identification of users. Bank generates mCard ID and password against user account. mCard requires user ID and password every time when the application is lunched. The flows of operation of mCard are:-

- Get user ID and Password

- Send user ID, Password and IMEI to Payment Gateway
- Verification of user
- Established connection between mCard and Payment Gateway

Figure 4 shows the flows of operation of mCard.

C. Operation of mCard at point of sale

POS merchants are already connected to the payment gateway (NPS). All the merchants have their own QR-code [10] (Quick Response Code) provided by NPS. The QR-code is used for identification of merchants. The figure 5 shows the details operation of mCard in POS terminal.

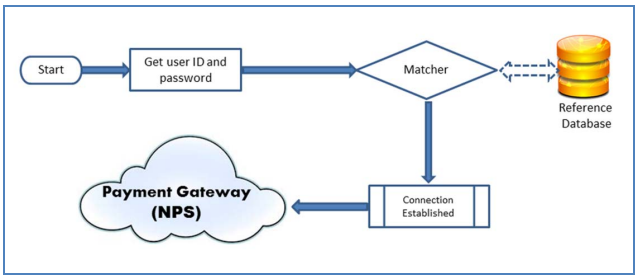


Figure 5: Flows of operation of mCard at POS terminal

D. Implementation

Due to the shortest of time the proposed model is not totally completed. The hardware requirements are Android [8] based Smart Phone or Tabs. Eclipse IDE [9] with Android Development Tools will be used for developing the mCard application.

IV. Security

Today, mobile applications security is very important taking in account the high number of mobile devices and users. World-wide, about one million people were infected by mobile malware in the first semester of 2011 [12]. According to [13] only 4% of smart-phones and tablets use anti-malware and anti-virus software and in the same period the number of mobile applications containing malware on marketplaces grew from 80 to 400 [14]. In [15] is presented a history of the most important mobile malware and several tools that can be used for malware analysis. The market share of Android based devices is increasing very fast and due to its open source system many attackers are focusing on it.

A. Mobile Applications Security Issues

Mobile devices are vulnerable to attacks as any other computers even the number of attacks is still reduced compared with personal computers.

The following technologies that exist on many mobile devices can be used by attackers to send and/or receive confidential data, malicious applications and to make unauthorized actions:

- Bluetooth – applications use it to connect to other devices
- Telephony – unauthorized phone calls are made, resulting in high costs or unauthorized recordings
- Messaging (Short or Multimedia Messaging Services) – used to send confidential content to attacker or to access paid numbers
- Wireless networks – used to connect to Internet to transfer data
- NFC (Near Field Communication) – used for unauthorized payments

Data services are used in order to transfer from and to malicious applications. The Internet access allows sending/receiving e-mail messages and unauthorized access to Internet resources (Web sites, Web services, Internet servers etc.)

The behavior of malicious applications could vary from annoying messages to very unrecoverable damages. The possible actions are:

- annoying messages
- unwanted web pages opened
- advertising popup
- higher costs resulting from messages sent (SMS, MMS), phone calls and payments
- unauthorized use of personal information data
- confidential data transferred on a remote location
- Altering data stored in file system, contacts, messages etc.

Depending on the platform and/or operating system, applications can

- read and write contacts
- read and write calendar entries
- read short/multimedia messages
- send e-mail messages, SMS and MMS
- make phone calls
- get location
- access the internet
- Read and write on file system.

Not all operating systems and platform allow all of these actions that can be dangerous if they are used by a malicious application. Also the user is not always informed about these actions when install the application.

B. Android Applications Security Model

Android operating system is based on *Linux kernel* so that applications isolation, the file system and security rules are *Linux* specific.

As can be seen from Figure 6, users can interact with Android device using *Linux shell* provided through *adb tool*. Usually development devices have root access granted with full rights, but this access is not available on most of the Android devices. Android applications files are packed in files with *apk* extension. These files contain all the classes and resources required by the applications [16]. The binary classes are converted into a proprietary format for *Dalvik* Virtual Machine (*dex files*). Usually, when applications are launched for the first time, the *dex* files are optimized and the resulting files are stored in /data/dalvik-cache directory. When the same

application is launched again, the optimized code is loaded from that special directory.

```
c:\android-sdk-windows\platform-tools>adb -d shell
$ ls -l
ls -l
drwxr-xr-x root     system            2011-10-03 15:44 app-cache
drwx------ root     root              2011-10-03 15:44 config
lrwxrwxrwx root     root              2011-10-03 15:44 sdcard -> /mnt/sdcard
drwxr-xr-x root     root              2011-10-03 15:44 acct
drwxrwxr-x root     system            2011-10-03 15:44 mnt
lrwxrwxrwx root     root              2011-10-03 15:44 etc -> /system/etc
drwxrwx--x system   system            2011-10-03 15:44 vendor
drwx------ root     root              2011-10-04 13:26 devlog
drwxrwx--- system   cache             2011-10-04 13:29 cache
-rw-r--r-- root     root         4429 1970-01-01 02:00 ueventd.rc
-rw-r--r-- root     root          758 1970-01-01 02:00 ueventd.pyramid.rc
-rw-r--r-- root     root            0 1970-01-01 02:00 ueventd.goldfish.rc
drwxr-xr-x root     root              2011-08-13 12:28 system
drwxr-xr-x root     root              2011-10-03 15:44 sys
drwxr-x--- root     root              1970-01-01 02:00 sbin
dr-xr-xr-x root     root              1970-01-01 02:00 proc
-rwxr-x--- root     root        20380 1970-01-01 02:00 init.rc
-rwxr-x--- root     root         9328 1970-01-01 02:00 init.pyramid.rc
-rwxr-x--- root     root         1677 1970-01-01 02:00 init.goldfish.rc
-rwxr-x--- root     root       107208 1970-01-01 02:00 init
-rw-r--r-- root     root          118 1970-01-01 02:00 default.prop
```

Figure 6: Android command line shell

Each application runs in a separate virtual machine, having its own unique identifier (UID) Figure 7. Due to this, application resources are protected from other applications and the communication and data transfer between applications has a high degree of confidence.

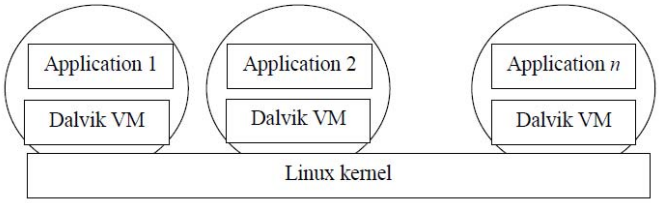


Figure 7: Android applications

V. COMPARISON

The Table 1 shows the comparison among Mobile banking, mCard and traditional plastic card.

Table 1: Comparison between mCard and existing system

	Traditional Card	**mCard**	**Mobile Banking**
Physical Damage	Yes	N/A	N/A
Economically Feasible	No	Yes	require extra charge
Eco-friendly	No (use Plastic)	Yes	Yes
Transaction fee	free	free	Yes
Merchants/ATM	Available	Available	Limited
Lost/Stolen	Yes	N/A	N/A
Replacement fee	Yes	Nil	N/A
Carry	Carry the Card	Only carry the phone	Only carry the phone

VI. CONCLUSION

The main discussion of the paper is "mCard" a mobile payment System, which replaced the traditional plastic based Credit/Debit/ATM cards. The system is under developing process. The system is developing for Android enable smartphones/ Tabs. In future the system is extends for iPhones and other mobile phones.

ACKNOWLEDGMENT

This paper is made possible through the help and support from everyone, including: parents, teachers, family, and friends.

Especially, please allow us to dedicate our acknowledgment to ***Dr. Shamim H Ripon***, Associate Professor, Department of CSE, East West University, Bangladesh for his most support and encouragement.

REFERENCES

[1] Telecompaper, "Bangladesh nearing 100 million mobile subs," February 02, 2013.

[2] Y.A. Au & R.J. Kauffman, "The economics of mobile payments: Understanding stakeholder issues for an emerging financial technology application," Electronic Commerce Research and Applications, 2007.

[3] S. Karnouskos & F. Fokus, "Mobile Payment: a journey through existing procedures and standardization initiatives," IEEE Communications Surveys and Tutorials, 2004.

[4] T. Dahlberg et al, "Past, present and future of mobile payments research: A literature review," Electronic Commerce Research and Applications, 2007.

[5] J. Ondrus & Y. Pigneur, "An Assessment of NFC for Future Mobile Payment Systems," International Conference on the Management of Mobile Business, 2007.

[6] PRWEB, "BRAC Bank's bKash," July 22, 2011.

[7] The Daily Star, "Bangladesh Bank opens gateway to e-commerce," December 28, 2012.

[8] Jason Chen, "An Introduction to Android," 2008.

[9] Mark Dexter, "Eclipse And Java For Total Beginners Companion Tutorial Document".

[10] "QR Code," [Online]. Available at:- http://www.qrcode.com.

[11] "International Mobile Station Equipment Identity," [Online]. Available at: - http//www.wikipedia.com.

[12] The Associated Press, Smartphone malware infections growing fast [Online]. Available at: http://bulawayo24.com, August 2011.

[13] J. Leyden, "Mobile app malware menace grows," [Online]. Available at: http://www.theregister.co.uk, August 2011.

[14] J. Scott, "Mobile devices in danger of attack, [Online]. Available at: http://www.itpro.co.uk, August 2011.

[15] H. Dwivedi, C. Clark and D. Thiel, "Mobile Application Security," New York: McGraw-Hill, 2010, pp. 364-389.

[16] E. Burnette, "Hello, Android: Introducing Google's Mobile Development Platform," 3rd Edition. The Pragmatic Bookshelf, 2010,p.33.

Structural Similarity Based Disparity Estimation for Stereoscopic Images

Mohammad Minhaz Akram
Dept. Of Electrical & Electronics Engineering
Ahsanullah University of Science & Technology
Dhaka-1208, Bangladesh
eminhaz@gmail.com

Hajera Siddiqa, Z. M. Parvez Sazzad
Dept. of Applied Physics, Electronics and Communication Engineering
University of Dhaka,
Dhaka-1000, Bangladesh.
hajerasiddiqa@yahoo.com.au, sazzad@univdhaka.edu

***Abstract*— Depth estimation of a scene from stereo images is an active area of research in computer vision applications. In this paper, a simpler and faster technique for depth estimation, based on structural similarity index (SSIM) method, is proposed. The technique provides a simple block based similarity measure that considers the structural similarity between the corresponding blocks of the left and right views. It avoids explicit calculations in depth or disparity estimation. The method shows that the estimation performance is acceptable and it can be used in autonomous navigation and 3D media applications.**

***Index Terms*— Depth estimation, SSIM, disparity, similarity measure, quality assessment.**

I. Introduction

Stereo depth estimation has the capability of estimating the distances of the scene's objects from the observer that looks at the scene from two slightly different vantage points. It has important implications in many real-world visual application domains, such as in entertainment (video, games), education [1] and medical application like body exploration [2], [3], therapeutic purposes [4], and so forth. The depth can be estimated by comparing the image pairs taken from the left and right views of a stereo system [5]. A number of methods have been proposed for stereo disparity measurement over the last two decades and a comprehensive analysis on stereo matching can be found in [6]. Amongst them block matching method has been attracted to find the disparity because of its better performance. Three classes of metrics are commonly used for block matching: correlation [7], [8], intensity differences [9], [7] and rank metrics [8], [6]. Two simple and widely employed similarity measures are sum of absolute and square differences (SAD and SSD, respectively) which are used in many real time applications [9], [7]. Eslami et al. [10] developed a new block based method which depends on finite ridge let transform to employ different weights for smooth and high contrast area of an image.

In this work, we propose a simple block based similarity measure that considers the structural similarity [11] between the corresponding blocks of the left and right views which has not been considered in conventional block matching disparity estimation. We provide this approach details in the following Section. The rest of this paper is organized as follows. In Section 2, we discuss the details of our disparity estimation. The result and analysis are presented in Section 3 and Section 4 concludes the paper.

II. Disparity Estimation

During disparity estimation, for each pixel in one image the corresponding pixel in the other image is sought, so that the corresponding pixels are the projections of the same 3D position. Afterwards, if the camera calibration is known, a depth map can be calculated from the disparity field. If a standard stereo setup is used, the corresponding pixels are constrained to lie on the same row. Thus, the search range for the disparity is 1-dimensional. As shown in Figure 1, a stereo vision that use to extract features from an image and locate the corresponding features in a second image to perceive depth of the captured scene. The relative pixel shifts from the left to the right image called disparity, are inversely proportional to the physical depth of the corresponding feature points in the scene. The disparity is measured horizontally and vertically for each matched pixel in the input images to produce a disparity map.

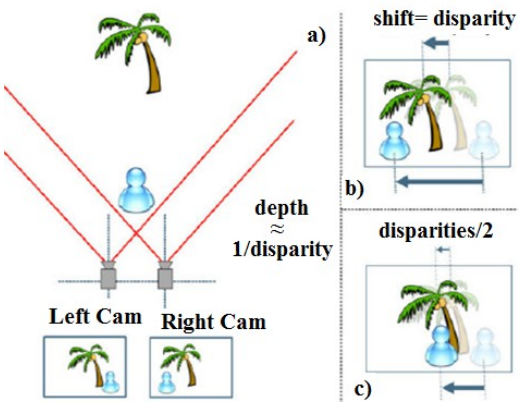


Fig. 1. Stereoscopic view in Computer Vision (a) the stereo view: left and right image, (b) the disparities by stereo matching, and (c) synthesized view by scaling disparities according to required viewpoint; image taken from [12]

The principle of the disparity estimation is to divide the left image into non overlapping 8×8 blocks. For each 8×8 block of

the left image, stereo correspondence searching is conducted based on maximum quality index and up to ±128 pixels of the right image (see Figure 2). As in the stereoscopic images database that we consider in our research is epi-polar rectified images, the displacement between the left and right view of a stereo pair is considered in horizontal direction only.

In order to measure disparity, firstly, the left image is segmented into non overlapping 8×8 blocks. Secondly quality index is measured in the two corresponding blocks between the left and right images. The quality index Q is defined by,

$$Q=\frac{4\sigma_{xy}\bar{x}\,\bar{y}}{(\sigma_x^2+\sigma_y^2)[(\bar{x})^2+(\bar{y})^2]} \tag{1}$$

where, $\mathbf{x}=\{x_i \mid i=1,2,\ldots,64\}$ and, $\boldsymbol{y}=\{y_i \mid i=1,2,\ldots,64\}$ be the left and right image blocks, respectively. And,

$$\bar{x}=\frac{1}{64}\sum_{i=1}^{64} x_i \quad , \quad \bar{y}=\frac{1}{64}\sum_{i=1}^{64} y_i \tag{2}$$

$$\sigma_x^2=\frac{1}{63}\sum_{i=1}^{64}\left(x_i-\bar{x}\right)^2 , \quad \sigma_y^2=\frac{1}{63}\sum_{i=1}^{64}\left(y_i-\bar{y}\right)^2 \tag{3}$$

$$\sigma_{xy}^2=\frac{1}{63}\sum_{i=1}^{64}\left(x_i-\bar{x}\right)\left(y_i-\bar{y}\right) \tag{4}$$

In these equations, $\bar{x}$, $\bar{y}$, σ_x^2, σ_y^2, and σ_{xy}^2 are the estimates of the mean of x, mean of y, the variance of x, the variance of y, and the covariance of x and y.

The dynamic range of Q is [-1,1]. The best value 1 is achieved if and only if $x_i=y_i$ for all i=1,2,…….64. For each 8x8 block of the left image, disparity index (DI) is found by searching the position of maximum quality index up to ±128 pixels of the right image. The estimation approach is shown in Figure 3.

$$DI=\max\left\{\begin{matrix} Q_{-128}, Q_{-127}, Q_{-126}, \ldots\ldots\ldots\ldots \\ \ldots, Q_{-1}, Q_0, Q_1 \ldots\ldots\ldots\ldots\ldots\ldots, Q_{127}, Q_{128} \end{matrix}\right\}$$

Finally, the depth map can be found by taking the displacement of absolute normalized value of pixel position.

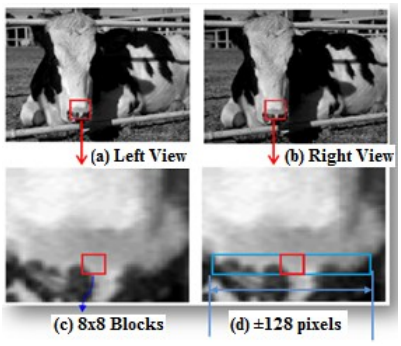


Fig. 2. (a) left view; (b) Right view ; (c)-(d) expanded view of (a) and (b); Given a pixel in the left image (c), a search window is defined in the right image (d). The best matched pixel is found by maximum disparity index within the window.

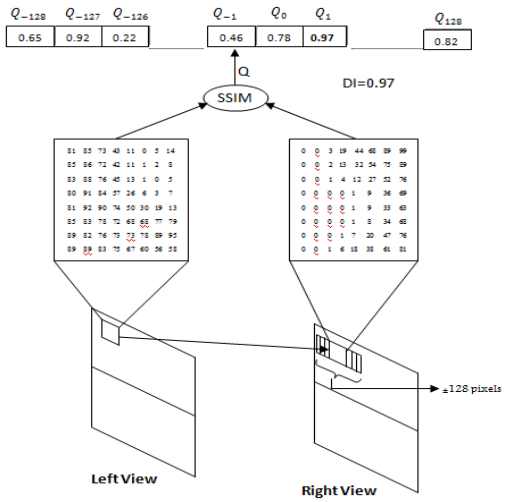


Fig. 3. Disparity Estimation Approach.

III. Optimum Block Size and Search Area Detection of the Disparity Estimation Algorithm

In order to identify the optimum block-size and search area for the block- based disparity estimation algorithm, we consider symmetric stereo image pairs from MICT stereoscopic image database [13]. There are seventy symmetric pairs with reference and coded. Out of all, ten are reference stereo pairs and sixty are JPEG coded pairs. The seven quality scales (QS: 10, 15, 27, 37, 55, 79, and reference) are selected for the JPEG coder. Subsequently, we compare the effect of different block sizes and search areas on the disparity estimation. Here, three different block sizes, 4×4, 8×8, and 16×16, and four different search areas, ±32 pixels, ±64 pixels, ±128 pixels, and ±160 pixels are used for the evaluation. The performance is evaluated empirically and the evaluation results are shown in Table 1. The numerical values in Table 1 indicate that how many image pairs are sufficiently estimated out of 70 pairs by the proposed algorithm. Here, we use two reference pairs and its coded versions, Saboten and Car images, to verify the result. The disparity maps of the reference pairs for all the above mentioned block and search areas are shown in Figures 4, and 5. The colors in the disparity maps are indicated by vertical bars in right side for each of the figures which are the estimated depths of the image pairs. Here, the suitable block size and search area are estimated empirically. The figures (Figure 4, and 5) show that the performance of the disparity algorithm is sufficient for the block size 8×8 with search area of ±128 pixels..

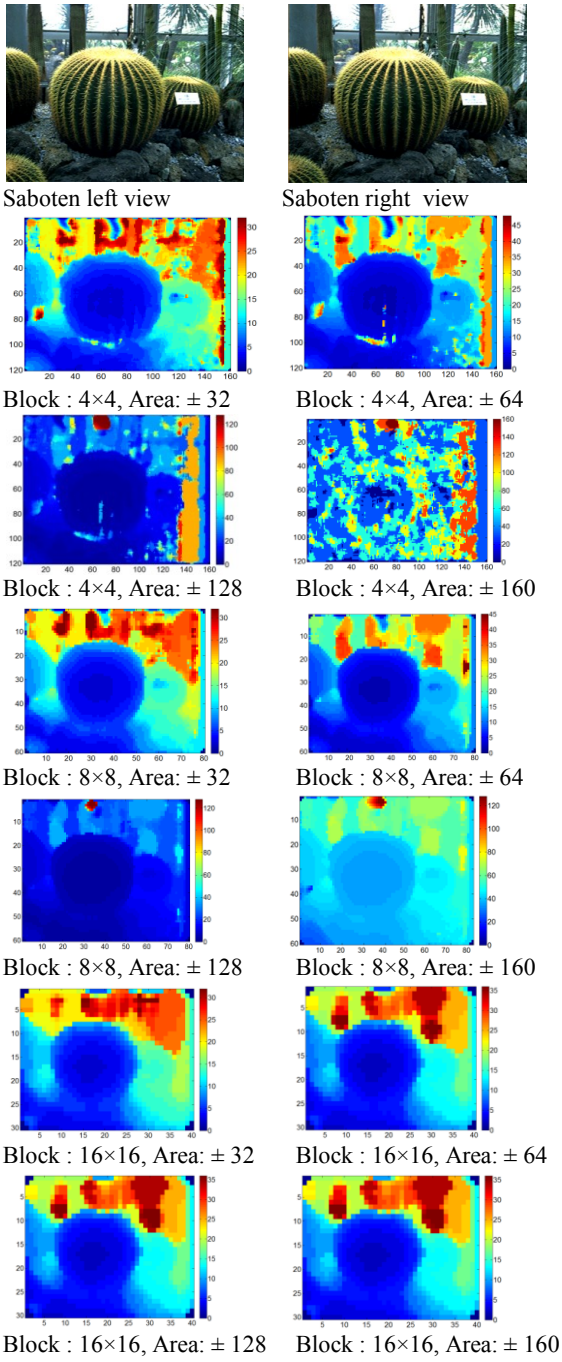


Fig. 4. Reference stereo pair (Saboten image) and its depth map for different block sizes and searching areas.

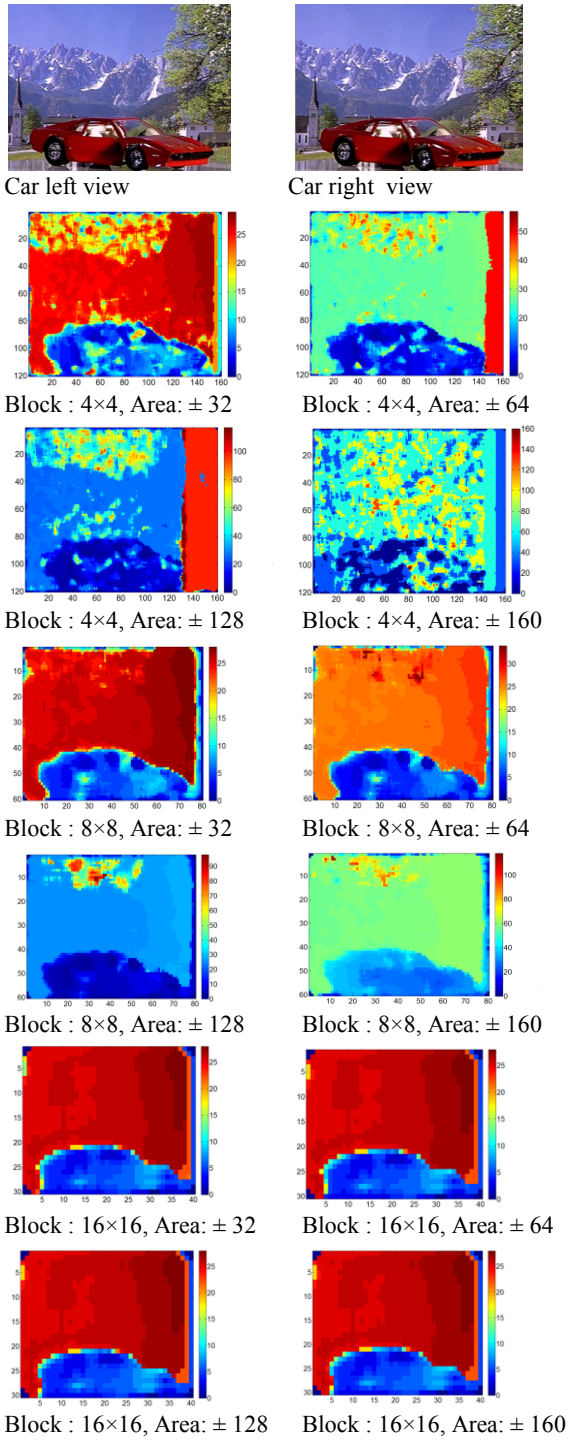
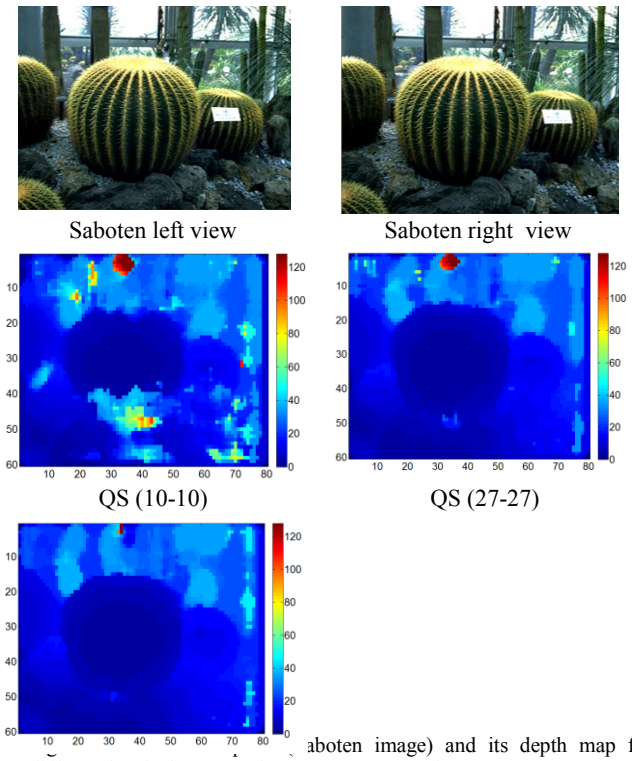


Fig. 5. Reference stereo pair (Car image) and its depth map for different block sizes and searching areas.

IV. Results and Discussion

Table 1 shows that the proposed algorithm can efficiently estimate disparity of 67 stereo pairs out of 70 pairs. Therefore, the disparity estimation performance of our proposed algorithm is all right. Consequently, the disparity maps of the coded pairs, considering different level of compression (Quality Scale, QS: 10, 27, and 55), are also shown in Figures 6, and 7 respectively for Saboten and Car image pairs. Figures 6, and 7 show that the performance of disparity estimation is degraded with increasing compression levels (i.e., decreasing in bit rate) for symmetric coded pairs. Although the block-based disparity algorithm is simple but the problem arises to localize the disparities accurately. The large search space increases the probability of sufficient intensity variation but introduces ambiguities and mismatches. Therefore, it reduces the selectivity of finding correct disparities, especially in the areas with depth discontinuities. Nevertheless, the structural similarity based measure may be a suitable measure in real-time disparity estimation. On the other hand, the algorithm in unstructured regions (e.g., uniform areas) may lead to ambiguity and mismatches because the same region in the left image will match equally well with many other similar blocks in the right image.

.



aboten image) and its depth map for different level of compressions (QS: 10, 27, and 55).

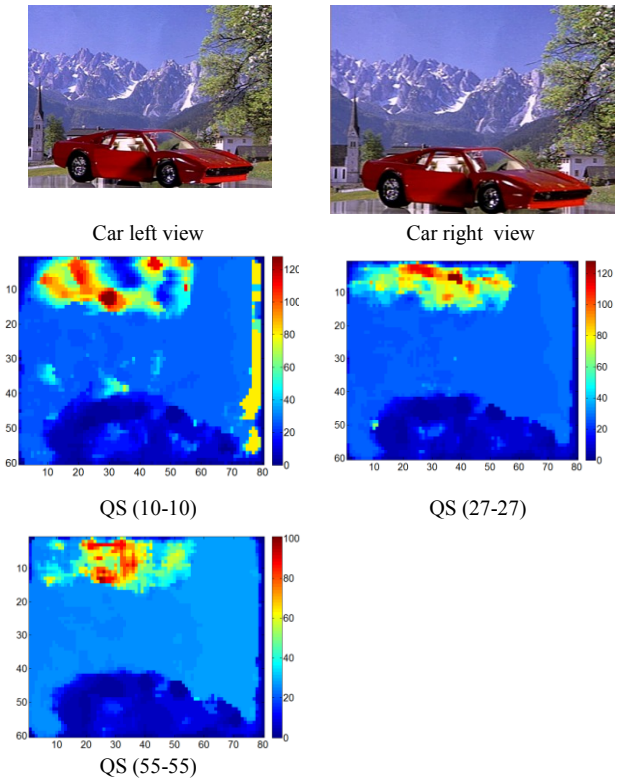


Fig. 7. Coded stereo pairs (Car image) and its depth map for different level of compressions (QS: 10, 27, and 55).

Optimum block size and area detection based on depth estimation performance

Block size	Searching area (Total no. of symmetric pairs = 70)			
	± 32 pixels	*± 64 pixels*	*± 128 pixels*	*± 160 pixels*
4×4	20	25	30	26
8×8	30	50	67	66
16×16	23	27	45	40

V. Conclusion

In this paper, we have introduced a structural similarity index based approach to reconstruct a disparity field. This allows simpler calculations in disparity estimation and is suitable for real time disparity estimation in different computer vision applications. We have considered the block size as 8x8 due to the JPEG compression. Though, the searching area of ±128 pixels is considered, this area can be defined based on the dimension specifically the width of the image. In this stydy the dimension of the images are 640x480. The width is 640 and searching area is ±128 i.e.less than ±($^1/_4$) of the width of the image. Hence, we can consider the searching area size as ±($^1/_4$) of the width of the image irrespective of the dimension and format of the image. In future, the algorithm can be extended for non epi-polar rectified images to evaluate exact objects' depth in meter that are suitable for robot navigation. Another

direction of future work is to extend the method that can be applied for any other coded images irrespective or certain coding techniques.

REFERENCES

[1] A. M. William and D. L. Bailey, "Stereoscopic visualization of scientific and medical content," in Proc. of the 33rd International Conference and Exhibition on Computer Graphics and Interactive Techniques (SIGGRAPH '06), p. 26, ACM, Boston, Mass, USA, July-August 2006.

[2] P. Ljung, C. Winskog, A. Persson, C. Lundstrom, and A.Ynnerman, "Forensic virtual autopsies by direct volume rendering," IEEE Signal Processing Magazine, vol. 24, no. 6, pp. 112–116, 2007.

[3] C.-F. Westin, "Extracting brain connectivity from diffusion MRI," IEEE Signal Processing Magazine, vol. 24, no. 6, pp. 124–152, 2007.

[4] Y. A. W. De Kort and W. A. Ijsselsteijn, "Reality check: the role of realism in stress reduction using media technology," Cyber psychology & Behavior, vol. 9, no. 2, pp. 230–233, 2006.

[5] M.R.M. Jenkin, A.D. Jepson, and K. Tsotsos, "Techniques for disparity measurement," CVGIP: image understanding, 53:14-30, 1991.

[6] D. Scharstein, and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," Int. J. of Comp. vision, vol. 47(1-3):7-32, Apr. 2002.

[7] Brown MZ., Burschka D., and Hager GD., "Advances in computational stereo," IEEE Trans. on PAMI 25:993-1008, 2003.

[8] Pan WH., Wei S., and Shang-Hong L., "A hybrid motion estimation approach based on normalized cross correlation for video compression," IEEE Int Conf ICASSP 2008:1037- 40.

[9] Wang Y., Ostermann J., and Zhang YQ., "Video processing and communication," Engle-wood Cli_s, NJ: Prentice Hall; 2001 pp. 1918-1927, Nov. 2005.

[10] M. Eslami, and F. Torkamani-Azar, "Block-based disparity estimation by partial finite ridgelet distortion search (PFRDS)," Accepted in Optics and Lasers in Engineering (on line).

[11] Z. Wang, A.C. Bovik, H.R. Sheikh and E.P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," IEEE Trans Image Process., vol. 13, no. 4, pp. 600-612, Apr. 2004.

[12] C.-F. Westin, \Extracting brain connectivity from di_usion MRI," IEEE Signal Processing Magazine, vol. 24, no. 6, pp. 124-152, 2007.

[13] Media Information and Communication Technology (MICT) Laboratory, University of Toyama, MICT stereo image database, http://mict.eng.u-toyama.ac.jp/mict/index2.html.

Bangla Phonetic Feature Table Construction for Automatic Speech Recognition

Foyzul Hassan, Mohammed Rokibul Alam Kotwal and Mohammad Nurul Huda
Department of Computer Science and Engineering
United International University
Dhaka, Bangladesh
foyzul.hassan@gmail.com, rokib_kotwal@yahoo.com, mnh@cse.uiu.ac.bd

***Abstract*—This This research constructs a phonetic feature (PF) table for all the phonemes pronounced in Bangla (widely known as Bengali) language where the whole study is divided into two parts. In the first part, a PF table is constructed, while the second part deals with Bangla automatic speech recognition (ASR) using PFs. For Bangla language, fifty three phonemes including both vowels and consonants are considered in which the phones, শ (/s/) and স (/s/), and, ণ (/n/) and ন (/n/) contain approximately same spectrum and hence, they share same PFs. In the PF table, twenty two PFs (Silence, Short Silence, Stop, …) are required for representing all the Bangla phonemes. On the other hand, the second part comprised of three stages: i) first stage deals with acoustic features, mel frequency cepstral coefficients (MFCCs) extraction, ii) second stage embeds PFs extraction procedure using a multilayer neural network (MLN) and iii) the final stage integrates a triphone-based hidden Markov model (HMM) for generating the output text strings by inputting log values of twenty two dimensional PFs. In the experiments on Bangla Newspaper Article Sentences, it is observed that the PF-based ASR system provides higher word correct rate, word accuracy and sentence correct rate in comparison with the standard MFCC-based method.**

Keywords— phonetic feature; mel frequency cepstral coefficient; multilayer neural network; automatic speech recognition; hidden Markov model

I. INTRODUCTION

There have been many literatures in automatic speech recognition (ASR) systems for almost all the major languages in the world. Unfortunately, only a very few works have been done in ASR for Bangla (can also be termed as Bengali), which is one of the largely spoken languages in the world. More than 220 million people speak in Bangla as their native language. It is ranked seventh based on the number of speakers [1]. A major difficulty to research in Bangla ASR is the lack of proper speech corpus. Some efforts are made to develop Bangla speech corpus to build a Bangla text to speech system [2]. However, this effort is a part of developing speech databases for Indian Languages, where Bangla is one of the parts and it is spoken in the eastern area of India (West Bengal and Kolkata as its capital). But most of the natives of Bangla (more than two thirds) reside in Bangladesh, where it is the official language. Although the written characters of standard Bangla in both the countries are same, there are some sounds that are produced variably in different pronunciations of standard Bangla, in addition to the myriad of phonological variations in non-standard dialects [3]. Therefore, there is a need to do research on the main stream of Bangla, which is spoken in Bangladesh, ASR.

Some developments on Bangla speech processing or Bangla ASR can be found in [4]-[11], where various hidden Markov model (HMM)-based ASR systems have been developed. Most of these ASR systems make use of a preprocessed form, such as mel-frequency cepstral coefficients (MFCCs), of the speech signal, which encodes the time-frequency distribution of signal energy. However, these MFCC-based systems do not provide better recognition performance in real acoustic conditions (See Figure 1(a)). On the other hand, a system based on articulatory features (AFs) or phonetic features (PFs) exhibits higher recognition accuracy in practical conditions and models coarticulatory phenomena more naturally [12](See Figure 1(b)). From the Figures 1(a) and 1(b), it is shown that the PF-based system outputs few misclassifications. The main problem for the Bangla language is that PF table is yet to be constructed.

In this paper, we have designed a phonetic feature (PF) table for all the phonemes pronounced in Bangla language. The first part of the research deals with a PF table construction, while the second part constructs a Bangla ASR using PFs. In the PF table, twenty two PFs are required for representing all the Bangla phonemes. On the other hand, the second part comprised of three stages: i) first stage deals with acoustic features, mel frequency cepstral coefficients (MFCCs), extraction, ii) second stage embeds PFs extraction procedure using a multilayer neural network (MLN) and iii) the final stage integrates a triphone-based HMM for generating the output text strings by inputting log values of twenty two dimensional PFs.

The paper is organized as follows. Section II briefly describes an approximate phonetic scheme and speech corpus for Bangla and formation of words, and speech corpus for Bangla. Section III explains about Bangla PFs, while Section IV deals with Proposed ASR construction using Bangla PFs. Again, Section V gives experimental setup, results and discussion on Bangla continuous word recognition. Finally, Section VI draws some conclusions with future directions.

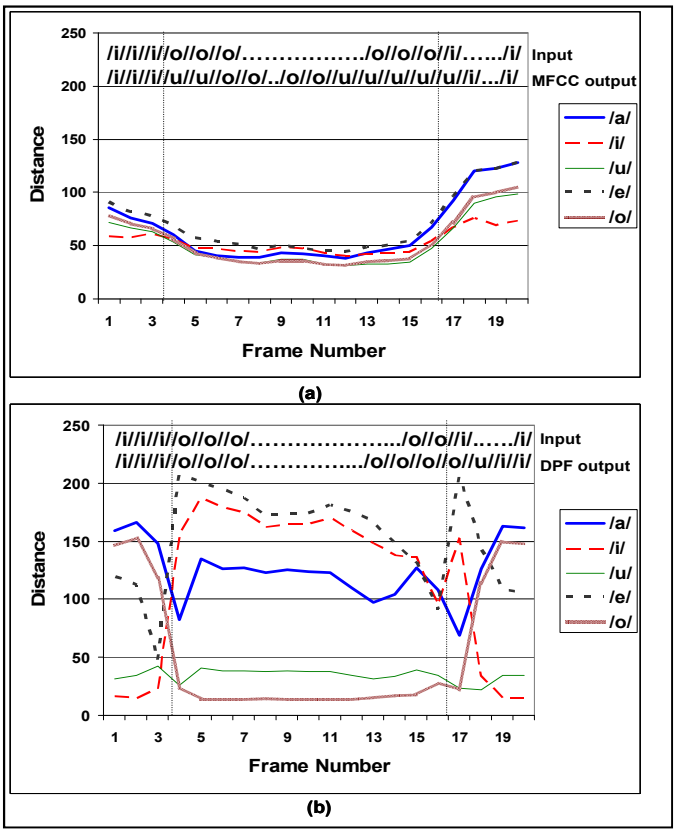


Fig. 1. Phoneme distances for utterance, /ioi/ using (a) MFCC-based system and (b) PF-based system.

II. Phonetic Scheme and Corpus for Bangla

A. Bangla Phonemes

The phonetic inventory of Bangla consists of 14 vowels, including seven nasalized vowels, and 29 consonants. An approximate phonetic scheme in IPA is given in [13][14], where only the main 7 vowel sounds are shown, though there exists two more long counterpart of /i/ and /u/, denoted as /i:/ and /u:/, respectively. These two long vowels are seldom pronounced differently than their short counterparts in modern Bangla. There is controversy on the number of Bangla consonants.

B. Bangla Words

TABLE I. Examples of Some Bangla Words With Their IPA

Bangla Word	English Pronunciation	IPA	Our Symbol
আমরা	AAMRA	/a m r a/	/aa m r ax/
আচরণ	AACHORON	/a tʃ r n/	/aa ch ow r aa n/
আবেদন	ABEDON	/a b æ d̪ n/	/ax b ae d aa n/

Table I lists some Bangla words with their written forms and the corresponding IPA. From the table, it is shown that the same 'আ' (/a/) has different pronunciation based on succeeding phonemes 'ম', 'চ' and 'ব'. These pronunciations are sometimes long or short. For long and short 'আ' we have used two different phonemes /aa/ and /ax/, respectively. Similarly, we have considered all variations of same phonemes and consequently, found total 51 phonemes excluding beginning and end silence (/sil/) and short pause (/sp/).

C. Bangla Speech Corpus

Hundred sentences from the Bengali newspaper "Prothom Alo" [15] are uttered by 30 male speakers of different regions of Bangladesh. These sentences (30x100) are used as training corpus (D1). On the other hand, different 100 sentences from the same newspaper uttered by 10 different male speakers are used as test corpus (D2). All of the speakers are Bangladeshi nationals and native speakers of Bangla. The age of the speakers ranges from 20 to 40 years. We have chosen the speakers from a wide area of Bangladesh: Dhaka (central region), Comilla – Noakhali (East region), Rajshahi (West region), Dinajpur – Rangpur (North-West region), Khulna (South-West region), Mymensingh and Sylhet (North-East region). Though all of them speak in standard Bangla, they are not free from their regional accent.

III. Proposed Bangla Phonetic Features

A phoneme can easily be identified by its PFs [16][17]. Table II specified in succeeding page of this paper shows the proposed Bangla PFs for all the phonemes with their international phonetic alphabet (IPA) and Bangla orthographic transcription. Here, the fifty three Bangla phonemes and twenty two PFs for each phoneme are silence, short silence, stop, nasal, bilabial, fricative, liquid, lenis, vowel, front, central, back, unvoiced, long, short, dipthong, high, low, medium, round, unround and glottal, which shown in the table horizontally and vertically, respectively. Here, (Front, Back, Central) and (High, Low, Medium) represent tongue position in forward and backward, and upward and downward directions, respectively. Besides, plus (+) and minus (-) elements in the table represent whether corresponding element is present or absent, respectively.

IV. Proposed ASR System using PFs

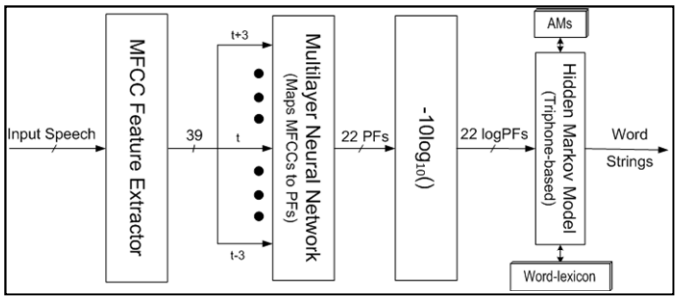


Fig. 2. Proposed PF-based ASR System.

We have implemented a PF-based ASR system with an input acoustic vector of MFCCs using an MLN which is shown in Figure 2. This system comprised of three stages: i) first stage deals with acoustic features, MFCCs extraction, ii) second stage embeds PFs extraction procedure using an MLN and iii) the final stage integrates a triphone-based HMM for generating the output text strings by inputting logarithmic values [17] of

TABLE II. PROPOSED BANGLA PHONETIC FEATURES

Bangla Orthographic Transcription	আ	চ	ছ		ও	র	প	, ন	গ	ঊ	ল	আ (short)	ম	ব	এ্যা		দ	ঐ	ক	র	ই	ধ	ড	প	শ/স	ও	ব	থ	ঙ	এ (long)	ভ	ঙ	স		হ	ক্ষ	ঘ	আ (long)	আ	চ	ম	য	থ	অ	হ	ই	ম	ঝ	জ	য (long)	য	ও (long)		ভ (long)	ঊ
IPA	/a/	/tʃ/	/tʃʰ/	Short pause	/o/	/ɾ/	/n/	/g/	/ɦ/	/l/	/ā/	/m/	/b/	/æ/	/d/	/i/	/k/	/ɾ/	/ē/	/dʰ/	/t̪/	/p/	/s/	/ɔ/	/s/	/tʰ/	/ŋ/	/e/	/bʰ/	/ŋ/	/ɛ/	/ɔ/	/u/	/h/	/pʰ/	/tʃʰ/	/ā/	/aː/	/dʰ/	/gʰ/	/dʒ/	/kʰ/	/ɾ̥/	/h/	/i/	/dʒ/	/dʰʒ/	/dʒ/	/ej/-	/ej/-	/ō/	Silence	/bʰ/	/u/
PFs/Phoneme	aa	ch	ey	sp	ow	r	n	g	uh	l	ax	m	b	ae	d	iy	k	er	ih	dh	t	p	sh	ao	s	th	ng	eh	v	nh	oy	uw	hh	f	chh	ay	ah	dha	gh	z	kh	aw	h	i	zh	jh	j	ky	y	w	sil	bh	u
Silence	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-
Short Silence	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Stop	-	-	-	-	-	-	-	+	-	-	-	-	+	-	+	-	+	-	-	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	+	-	-	-	+	+	+	-	-	-	-	+	-
Nasal	-	-	-	-	-	-	+	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Bilabial	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-
Fricative	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	+	-	+	+	-	-	+	-	-	-	-	+	+	-	-	-	-	+	-	-	+	-	+	+	+	-	-	-	-	-	-
Liquid	-	-	-	-	-	+	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	-	-	-
Lenis	-	-	-	-	-	-	-	+	-	-	-	-	+	-	+	-	-	-	-	+	-	-	+	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	+	+	-	-	-	-	+	-	-	-	-	-	-	-	-
Vowel	+	-	+	-	+	-	-	-	-	-	+	-	-	-	-	+	-	+	+	-	-	-	-	+	-	-	-	+	-	-	+	+	-	-	-	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+
Front	-	-	-	-	-	-	-	-	-	-	-	+	+	+	-	+	-	-	+	-	-	+	-	-	-	-	-	+	+	-	-	-	-	+	-	-	-	-	-	-	-	-	+	-	-	-	-	-	+	-	-	-	-
Central	+	-	-	-	-	+	+	-	-	+	-	-	-	-	+	-	-	+	-	+	+	-	+	+	+	+	-	+	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-
Back	+	+	-	-	-	-	-	+	+	-	+	-	-	-	-	-	+	-	-	-	-	-	+	-	-	-	+	-	-	-	-	+	+	-	+	-	-	-	+	-	+	-	-	-	-	+	-	+	+	-	-	-	+
Unvoiced	-	+	-	+	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	+	+	+	-	+	+	-	-	-	-	-	-	+	+	+	-	-	-	-	-	+	-	-	-	+	+	-	-	-	-	+	-	-
Long	+	-	-	-	+	-	-	-	-	+	-	+	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Short	+	-	+	-	-	-	-	-	+	-	+	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	+	-	-	+	-	-	-	-	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Dipthong	+	-	+	-	-	-	-	-	-	+	-	+	-	-	-	-	-	+	-	-	-	-	-	-	-	-	+	-	-	-	+	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-
High	+	-	-	-	-	-	-	-	-	-	+	-	-	-	-	+	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	+
Low	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	+	-	-	+	-	-	-	-	+	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-
Medium	+	-	+	-	-	-	-	-	-	+	+	+	-	+	-	-	-	+	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-
Round	+	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	+
UnRound	+	-	+	-	-	+	-	-	-	+	+	+	-	-	-	+	-	+	+	-	-	-	-	-	-	-	-	+	-	-	-	-	-	+	-	-	+	+	-	-	-	-	+	-	-	-	-	-	+	-	-	-	-
Glotal	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-

twenty two dimensional PFs. The thirty nine dimensional MFCCs extracted in the first stage are entered into the MLN with five layers including three hidden layers after combining a current frame x_t with the other two frames that are three points before and after the current frame (x_{t-3}, x_{t+3}) where the MLN generates twenty two PF values for each input frame of 39x3 features. The three hidden layers comprised of 400, 200 and 100 units, respectively. The MLN is trained using the standard back-propagation algorithm.

V. EXPERIMENTS

A. Setup

For evaluating word recognition performance, word correct rate (WCR), word accuracy (WA) and sentence correct rate (SCR) for D2 data set are evaluated using an HMM-based classifier. The D1 data set is used to design Bangla triphone HMMs with five states, three loops, and left-to-right models. Input features for the classifier are 39 dimensional MFCCs and log values of 22 dimensional PFs. The mixture components are set to 1, 2, 4 and 8.

For evaluating the performance of standard MFCC-based method including the proposed method, we have designed the following experiments:

(a) MFCC:dim-39 [Baseline]

(b) PF:dim-22 [Proposed]

In our experiments the range of output is from 0 to 1, where the non-linear function is a sigmoid, (1/(1+exp(-x))) for the hidden and output layers of MLN. For evaluating PF correct rate we have considered 0.20 as threshold to obtain better segmentation. Here, 0.20 is considered as threshold by observing the experimental results.

B. Result Analysis and Discussion

Segmentation for silence, short silence, stop, nasal, bilabial, fricative, liquid, lenis, vowel, front, central, back, unvoiced, long, short, dipthong, high, low, medium, round, unround and glottal PFs are depicted in Figures 3 and 4 for ideal and real cases for utterance, /prothom/. From both the figures, it is observed that segments of nasal, liquid, vowel and front are more precise (follows ideal line) in Figure 3, and unvoiced, long, dipthong, high, low, medium, unround and glottal exhibit better segments with respect to ideal segmentation in Figure 4. Again, Figure 5 shows correct rates for each of the PFs using the test utterances in D2 data set, where PF correct rates for the corresponding PFs are 97.83%, 52.88%, 75.15%, 75.88%, 64.30%, 84.68%, 49.20%, 84.67%%, 95.72%, 87.83%, 88.22%, 78.42%, 93.79%, 87.49%, 86.65%, 82.97%, 77.82%, 70.75%, 92.62%, 86.15%, 89.00%, and 100.00%, respectively.

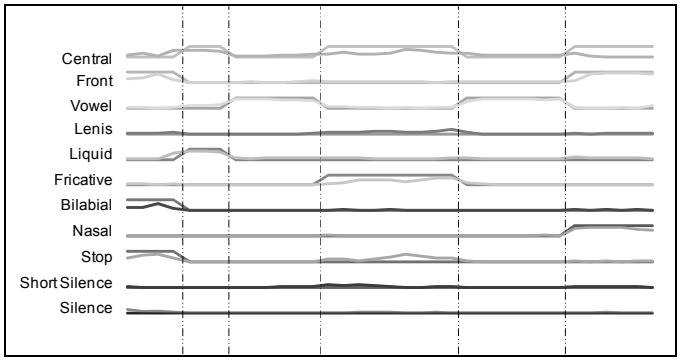


Fig. 3. Segmentation for silence, short silence, stop, nasal, bilabial, fricative, liquid, lenis, vowel, front and central PFs using the utterance /prothom/.

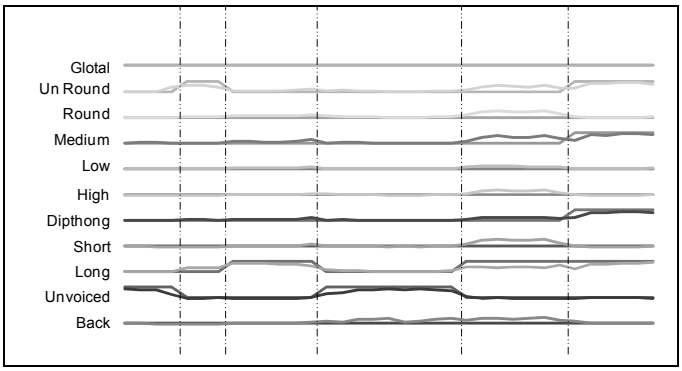


Fig. 4. Segmentation for back, unvoiced, long, short, dipthong, high, low, medium, round, unround and glottal PFs using the utterance /prothom/.

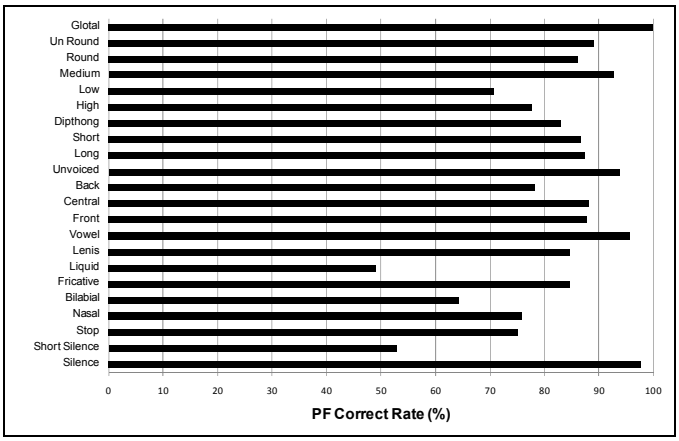


Fig. 5. Correct rates (%) for silence, short silence, stop, nasal, bilabial, fricative, liquid, lenis, vowel, front, central, back, unvoiced, long, short, dipthong, high, low, medium, round, unround and glottal PFs using the test utterances in D2 data set.

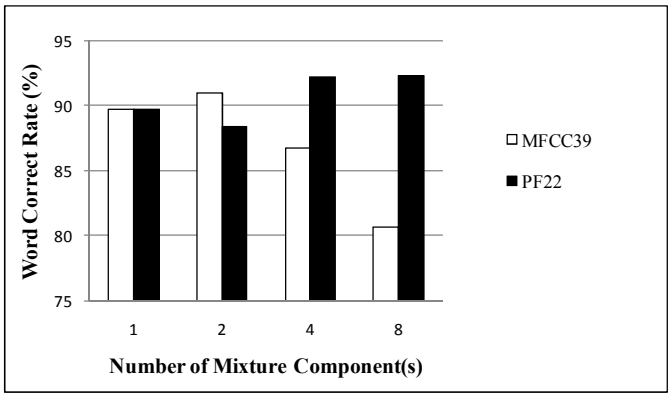


Fig. 6. Word Correct Rates for MFCCs and Proposed Method.

Figure 6 shows the comparison of word correct rates among all the investigated methods, standard MFCC-based method and proposed method. Among all the mixture components except two, the proposed method shows higher correctness in comparison with baseline. It is observed from the figure that the proposed method exhibits its best performance (92.25%) at mixture component eight. Besides, the mixture components, four and eight in the proposed method exhibit almost the same performance. Therefore, further investigation for higher correctness in higher mixture component is not required.

Word accuracies for the different investigated mixture components in standard MFCC-based and proposed methods are depicted in Figure 7. In mixture components one, two, four and eight, the proposed method provides 89.45%, 88.02%, 91.43% and 91.64% accuracies respectively, whereas 89.03%, 90.33%, 86.17% and 80.43% are observed in baseline method for the corresponding mixture components respectively.

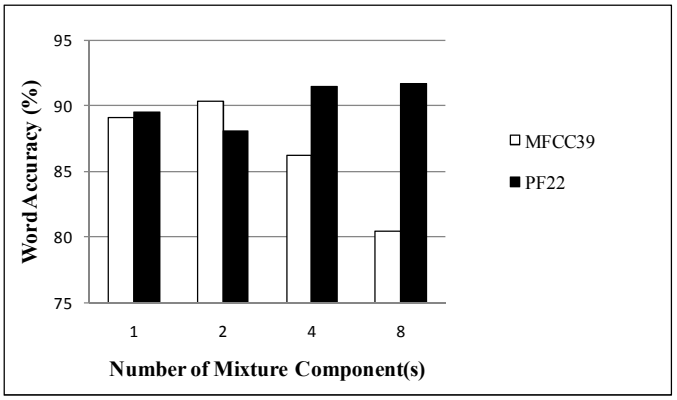


Fig. 7. Word Accuracies for MFCCs and Proposed Method.

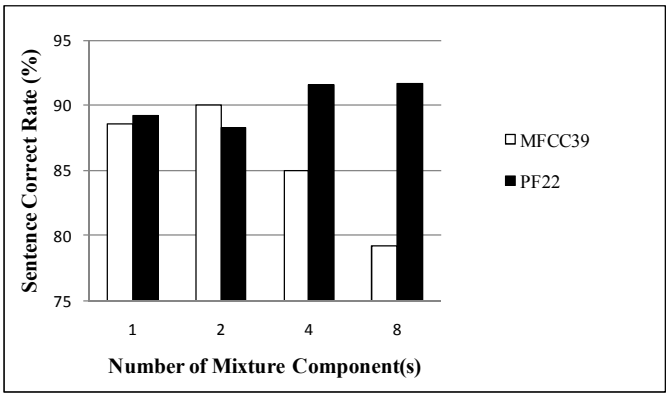


Fig. 8. Sentence Correct Rates for MFCCs and Proposed Method.

Sentence correct rate which is shown in Figure 8 gives an idea about the performance of ASR systems investigated. For the experimented mixture components, there are 89.20%, 88.20%, 91.50% and 91.60% SCRs are found in the proposed method respectively, while baseline system generates 88.60%, 90.00%, 85.00% and 79.20% for the same experimental conditions.

Table III exhibits word recognition performance with respect to correctly recognized words (H), deletion (D), substitution (S) and insertion (I), respectively for the experimented mixture components in both the investigated ASR systems using the 3290 input words. For proposed and baseline methods, the H, D, S and I are 3035, 52, 203 and 20 respectively; and 2654, 202, 434 and 8, respectively for the investigated mixture component eight. Here, proposed method inserted more words than baseline. On the other hand, sentence recognition information for the investigated mixture

components is provided in Table IV using 1000 input spoken sentences.

TABLE III. WORD INFORMATION FOR INVESTIGATED METHODS WHERE H, D, S AND I REPRESENT CORRECT WORDS, DELETION, SUBSTITUTION AND INSERTION OUT OF 3290 RESPECTIVELY

		Mixture Components			
		Mix 1	Mix 2	Mix 4	Mix 8
MFCC 39	H	2950	2992	2851	2654
	D	91	75	114	202
	S	249	223	325	434
	I	21	20	16	8
AF 22	H	2952	2908	3033	3035
	D	77	106	58	52
	S	261	276	199	203
	I	9	12	25	20

TABLE IV. SENTENCE INFORMATION FOR INVESTIGATED METHODS WHERE H, AND S REPRESENTS CORRECTLY AND INCORRECTLY RECOGNIZED SENTENCES RESPECTIVELY OUT OF 1000

		Mixture Components			
		Mix 1	Mix 2	Mix 4	Mix 8
MFCC 39	H	886	900	850	792
	S	1144	100	150	208
PF 22	H	892	882	915	916
	S	108	118	85	84

VI. CONCLUSION

This paper has constructed a phonetic feature table for Bangla automatic speech recognition. In the first part of the research twenty two phonetic features are considered for Bangla spoken language and the second part of the research designs an ASR system using the PFs considered here. The following conclusions are given:

(i) Segmentation for each of the PFs follows ideal boundaries for an input spoken sentence.

(ii) Correct rates for most of the PFs are above 80%.

(iii) Word correct rate, word accuracy and sentence correct rate for the proposed method using all the investigated mixture components except two are better in comparison with the standard MFCC-based method.

In near future, the author would like to evaluate PFs using recurrent neural network (RNN), which accommodates longer context window in its architecture. Moreover, the authors evaluate the experiments for gender independent environments.

REFERENCES

[1] http://en.wikipedia.org/wiki/List_of_languages_by_total_speakers, Last accessed April 11, 2009.

[2] S. P. Kishore, A. W. Black, R. Kumar, and Rajeev Sangal, "Experiments with unit selection speech databases for Indian languages," Carnegie Mellon University.

[3] http://en.wikipedia.org/wiki/Bengali_phonology, Last accessed April 11, 2009.

[4] S. A. Hossain, M. L. Rahman, and F. Ahmed, "Bangla vowel characterization based on analysis by synthesis," Proc. WASET, vol. 20, pp. 327-330, April 2007.

[5] M. A. Hasnat, J. Mowla, and Mumit Khan, " Isolated and Continuous Bangla Speech Recognition: Implementation Performance and application perspective, " in *Proc. International Symposium on Natural Language Processing* (*SNLP*), Hanoi, Vietnam, December 2007.

[6] R. Karim, M. S. Rahman, and M. Z Iqbal, "Recognition of spoken letters in Bangla," in *Proc. 5th International Conference on Computer and Information Technology (ICCIT02)*, Dhaka, Bangladesh, 2002.

[7] A. K. M. M. Houque, "Bengali segmented speech recognition system," Undergraduate thesis, BRAC University, Bangladesh, May 2006.

[8] K. Roy, D. Das, and M. G. Ali, "Development of the speech recognition system using artificial neural network," in *Proc. 5th International Conference on Computer and Information Technology (ICCIT02)*, Dhaka, Bangladesh, 2002.

[9] M. R. Hassan, B. Nath, and M. A. Bhuiyan, "Bengali phoneme recognition: a new approach," in *Proc. 6th International Conference on Computer and Information Technology (ICCIT03)*, Dhaka, Bangladesh, 2003.

[10] K. J. Rahman, M. A. Hossain, D. Das, T. Islam, and M. G. Ali, "Continuous bangle speech recognition system," in *Proc. 6th International Conference on Computer and Information Technology (ICCIT03)*, Dhaka, Bangladesh, 2003.

[11] S. A. Hossain, M. L. Rahman, F. Ahmed, and M. Dewan, "Bangla speech synthesis, analysis, and recognition: an overview," in *Proc. NCCPB,* Dhaka, 2004.

[12] K. Kirchhoff, et. al, "Combining acoustic and articulatory feature information for robust speech recognition," Speech Commun.,vol.37, pp.303-319, 2002.

[13] C. Masica, The Indo-Aryan Languages, Cambridge University Press, 1991.

[14] Ghulam Muhammad, Yousef A. Alotaibi and Mohammad Nurul Huda, "Automatic Speech Recognition for Bangla Digits," ICCIT'09, Dhaka, Bangladesh, December 2009.

[15] Daily Prothom Alo. Online: www.prothom-alo.com

[16] S. King and P. Taylor, "Detection of Phonological Features in Continuous Speech using Neural Networks," Computer Speech and Language 14(4), pp. 333-345, 2000.

[17] S. King, et. al, "Speech recognition via phonetically features syllables," Proc ICSLP'98, Sydney, Australia, 1998.

[18] T. Fukuda and T. Nitta, "Noise-robust ASR by Using Distinctive Phonetic Features Approximated with Logarithmic Normal Distribution of HMM," Proc. Eurospeech 2003, Vol.III, pp.2185-2188, Sep. 2003.

A Faster Approach to Periodic Data Flipping of SRAM Array for NBTI Recovery

Sani Md Ismail, Ismail Hossain
Dept. of Electrical, Electronic & Communication Engineering
MIST
Dhaka, Bangladesh

Md Shazzad Hossain
Dept. of Electrical & Electronic Engineering
AIUB
Dhaka, Bangladesh

Yeasir Arafat
Dept. of Electrical & Electronic Engineering
BUET
Dhaka, Bangladesh

***Abstract*—Negative Bias Temperature Instability (NBTI) is a prime reliability issue for micro and nano-scale semi-conductor devices. Due to continuous device scaling, NBTI effect has become more severe than before and affected device life-time. Memory devices like SRAMs are tremendously affected by NBTI as the PMOS transistor gate is connected at '0' logic for a long time. Several device-level and architecture-level solutions have been proposed to improve device life-time by interrupting NBTI degradation. Such an architectural level solution is to flip data in a particular SRAM cell after a certain time, causing periodic stress and relaxation. SRAM data flipping techniques proposed so far are not so time friendly as it is needed to access each memory cell individually and flip the data stored. It makes the process more time consuming and inconvenient for present ultra-fast system with high activity factor. In this paper, we proposed a new 7-T SRAM cell to allow flipping data of more than one memory cells at same clock pulse, hence decreasing the flipping time of entire memory array and concluded that with the new proposed cell and flipping procedure, data flipping of the entire memory array will become much faster which will ensure convenient NBTI recovery.**

Keywords—7-T SRAM Cell; Cell Data Flipping; Fast Flipping Time; NBTI Degradation; SRAM

I. INTRODUCTION

Device reliability is one of the prime concerns in recent VLSI design world. Integrated system design engineers face many issues and these are becoming more and more complicated with the continuous reduction in feature size. One of the prime issues is known as Negative Bias Temperature Instability which affects the transistor threshold voltage, operation delay, drive current etc., resulting complete device failure in a certain period of time. As NBTI is responsible for complete operational failure of the device, this issue is taken into account very seriously by present researchers.

Negative Bias Temperature Instability is known as generation of interface traps at Si-SiO_2 interface of MOS transistor under negative bias condition (V_{gs}= -V_{dd}) at elevated temperature. During oxidation, most of tetrahedral Si atoms bond to Oxygen and few atoms bond to Hydrogen, forming weak Si-H bonds. For PMOS transistor, when it is negatively biased (gate to source voltage), Si-H bonds are dissociated by holes in channel, creating dangling bonds which are called interface traps (Fig. 1). Interface traps are known as electrically active physical defects [1-2]. They are also called P_b centers [3]. Due to generation of interface traps, PMOS transistor faces several issues like increase in threshold voltage, mobility degradation, reduction in drain current and trans-conductance and increase in delay. After a certain period of time, mentioned parameters are so affected that device faces complete failure, becomes unable to operate anymore. Thus NBTI affects the life-time of electronic devices.

NBTI issue was known since 1966 [1, 4], but it was not so severe at early stage. At the beginning, BTI effect was quite ignored by the designers and the researchers. In today's nano electronic world, device area and size is a great concern that is why designers are always trying to minimize feature size as much as possible. From 10 µm process introduced in 1971, we are now living in the era of 22 nm process. Process technology is developing so rapidly minimizing feature size. It is proven in several researches that NBTI has become an alarming threat to circuit reliability and life-time in present nano-meter scale technology [2, 5-11].

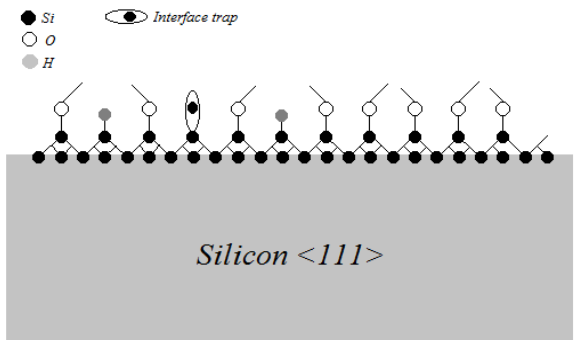


Fig. 1. Interface Trap in Silicon <111> substrate

Several nano-meter scale digital circuits are affected by NBTI, among which memory circuits like SRAM are most vulnerable to the issue as PMOS gates in the structure are connected to negative bias for comparatively longer time. Many solutions have been proposed to mitigate the effect of NBTI in memory cells. Architectural NBTI mitigation techniques are mainly focused on two sections: time-dependent stress-recovery techniques and power-control techniques. Power-control techniques includes power-gating [12-13], recovery boosting [14] etc. These techniques are focused on controlling the drive voltage in the memory array

in standby mode. Time-dependent stress-recovery techniques are focused on balancing the stress (negative bias) and recovery (positive bias) phase equally on both of the PMOS transistors in the memory cell, thus ensuring less interface traps generation compared to continuous stress after a certain period of time. It is observed that after a certain period of time, AC stress ensures less interface traps generation compared to continuous DC stress [1, 2, 15-22].

AC stress is mainly applied by flipping the data between the cross coupled inverters of a memory cell after a certain period of time. Flipping is done exactly at middle point of total stress-recovery time to ensure equal balance of stress and recovery at both PMOS transistor gates so that NBTI effect is mitigated equally at each gate. Data flipping frequency is defined as how many times the data in the entire memory array will be flipped in a certain amount of time (e.g. 24 hours). Flipping is only done in standby mode. Under no circumstances, flipping will be allowed while SRAM is in active mode (read /write mode) to ensure data integrity.

Although theoretically data flipping technique is a very effective process to mitigate NBTI effect, it is also a very time-consuming process to individually access to every cell of a memory array and flip each bit [2]. Data of a memory cell is generally flipped after each 12 hours a day. High frequency stress-recovery phase is avoided due to large flipping time of entire memory array. On the contrary, low frequency stress-recovery phase is highly susceptible to be affected by the activity factor of SRAM during active mode.

In this paper, we have proposed a new 7-T memory cell which is designed to flip more than one cell of an entire memory array at the same clock pulse. We have introduced one extra NMOS in the conventional 6-T memory cell called "flip MOS". The two cross-coupled inverters will exchange their complimentary bits among them with a pulse at the gate of this flip MOS. This proposed flipping process has ensured much less data flipping time of entire memory array compared to the previous techniques.

The rest of this paper is organized as: in section II, background of various data flipping techniques of SRAM have been discussed. In section III, the new 7-T SRAM cell has been proposed and the architecture of our proposed SRAM cell has been described in section IV. Consequently, in section V, we have discussed the time friendly characteristics of our design and in section VI, conclusion has been made on our finding.

II. Background of SRAM Data Flipping Technique

One of the most common architectural solutions for NBTI mitigation for SRAM is data flipping technique. As periodic stress-recovery phase in SRAM PMOS device provides less interface trap generation, data is flipped periodically in standby mode. Various methodologies of data flipping technique are described in [2]. There are mainly two approaches: software approach and hardware approach.

A. A Software Approach

In this methodology, data flipping is fully done by the processor. There is no extra hardware requirement inside SRAM device for flipping. The whole data flipping procedure is written as a procedural coding in the associated processor. At required time, processor interrupt the normal operation of SRAM, start reading the data from each cell, invert it and then put the inverted data back in that cell. Thus data from all locations (memory array address) is read by the processor and inverted data is put back to the corresponding address. There is no area overhead in this process. But it is obvious from the methodology description that it will be a very time consuming process as processor need to get access to each cell, read it, invert it and then put the inverted data back there. It is observed in [3] that the full flipping process for one cell requires around 200 clock cycles. If it is a 512×512 memory array, the total time for full flipping for the entire array will require around 6×10^7 clock cycles. For a 1 GHz system with 512×512 SRAM array, it will require significant time to complete the whole flipping process, which is not a very time friendly situation for today's fast data processing system. Moreover, the above mentioned minimum value may differ according to the distance of the SRAM block from the processor. Although this methodology provides area friendly solution, it is not a fruitful approach for fast design with high activity factor.

B. Hardware Approach

Hardware approach is a better solution than software approach regarding total flipping time although it introduces some extra area inside SRAM system. In this approach, almost all the procedures in flipping mode are performed inside SRAM system. So the extra time needed for the involvement of processor is not required here. But to perform the procedure inside SRAM block, it is needed to implement some logical hardware like multiplexers and registers inside SRAM. Each and every cell will be accessed, read, flipped and written back by an embedded control system inside SRAM control block. Total flipping time only depends on the size of the memory array, not on the distance from the processor. But due to implementation of some extra logic, memory access time will increase in active mode [2]. Although in case of data flipping period, it is much faster than the software approach methodology.

One thing is common for both methodologies that, the flipping time depends on number of cells present at the memory array. For the high density L3 caches, this flipping time is alarming for both approaches. Due to large flipping time requirements, flipping is done in a very low frequency rate (once a day). As NBTI mitigation is not so dependent on flipping frequency, experiments on different flipping frequencies were quite avoided. We have proposed a system where the total flipping time will be significantly less than any of the above mentioned two approaches. In the present work, a dynamic stress-recovery frequency approach has been proposed and observed that NBTI effect has mitigated significantly.

III. Proposed 7-T SRAM Cell

A 7-T SRAM cell is proposed with an extra control NMOS transistor called *flip MOS*. The schematic of the proposed memory cell is illustrated in Fig. 2.

In normal operation, the cell works as a normal 6-T SRAM cell. An extra control transistor *CMF* is connected with both *Bit* & *Bit_b*. In normal mode, *CMF* is deactivated, so *Bit* & *Bit_b* are insulated from each other. The cell works just like 6-T SRAM cell in this condition, performing read/write/ non-flipping standby etc. During data flipping, the *flip MOS CMF* is activated by providing a controlled positive pulse on the gate of *CMF* through Flip line (FL). When *CMF* is ON, *Bit* & *Bit_b* are connected. As these two nodes are preserving complementary logic (if one is holding 0 then surely another is holding 1), the high logic node will find a discharging path through the low logic node to ground. For example, if *Bit* is holding logic 1, then *Bit_b* is holding logic 0. When *CMF* will be on, logic 1 in *Bit* will be discharged through the path: *CMF-M4-gnd.* At the same time, *Bit* will gain some voltage. *CMF* is kept ON for a particular time so that *Bit* has discharged enough to work as low logic and *Bit_b* has gained enough to work as high logic. When *CMF* is turned off after controlled discharge, *Bit* will hold low logic (0) and *Bit_b* will hold high logic (1) in normal operation of two cross coupled inverters. Thus *Bit* & *Bit_b* have successfully exchanged logic among themselves, flipping the data in the memory cell. The same procedure is applicable if *Bit_b* is holding logic high and *Bit* is holding logic low. In that case, the discharging path will be: *CMF-M3-gnd.* So, by providing a controlled positive pulse, both cross-coupled inverters will exchange their logic among themselves.

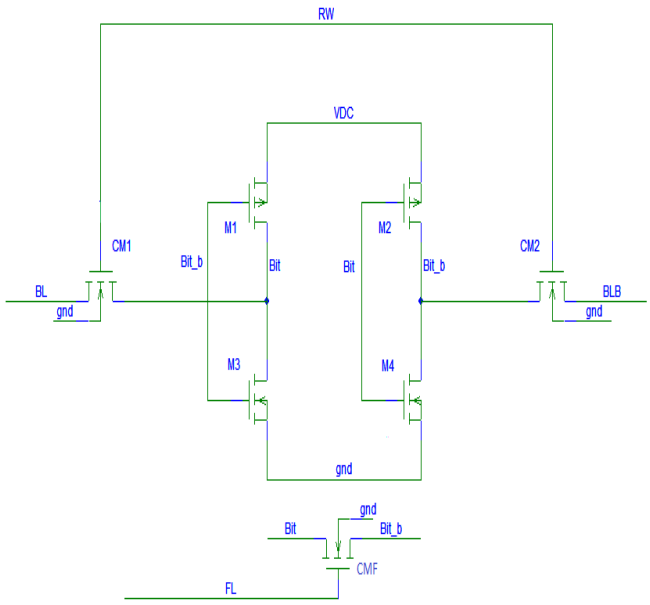


Fig. 2. Proposed 7-T SRAM Cell

It is very important to define the ON time of CMF during flipping. The pulse width of the flip pulse should be defined so accurately so that data is flipped successfully. If CMF is not on for enough time to perform adequate discharge of logic high node, data will not be flipped. Also, if CMF is kept ON for long period of time, then data may also not be flipped. Because, after adequate discharge, logic low node will turn into logic high node. If *CMF* is not turned OFF, the new logic high node may also start discharging throw the alternate discharging path, causing second time data flipping. In this case, data will flip but will go back to the previous state again. The flip data will not be stable. So, to maintain data integrity, it is very important to define the pulse width of the flip pulse properly.

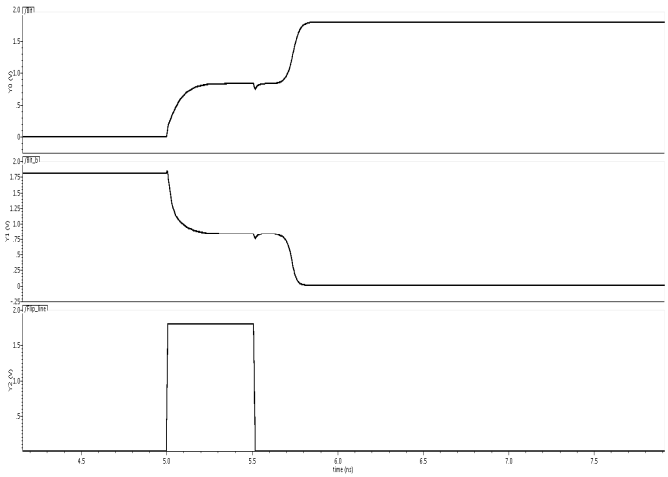


Fig. 3. Data Flipping in Proposed Memory Cell

The flip pulse width should be defined in such a way that two logic nodes will get enough time to exchange each other's logic and the exchanged logic will remain stable until any other action has been applied. We have run the simulation of our proposed cell in 180 nm process (Fig. 3). The memory cell is designed in *Cadence Virtuoso Schematic Editor* and simulation is done using *Spectre.*

High V_{th} MOSFET has been used as it is observed that high V_{th} are less affected by NBTI [13]. It has been observed the minimum and maximum value of flip pulse time for several W/L ratios for ensuring successful flipping. The range of flip pulse (maximum and minimum values) changes with different W/L ratios. Various flip pulse ranges with different W/L ratios are shown in Table I.

TABLE I: Flip Pulse Time for Different Transistor Width (180 nm Process)

Width (NMOS) (µm)	Width (PMOS) (µm)	Max Pulse Width (ps)	Min Pulse Width (ps)	Nominal Pulse Width (ps)	Nominal-Max/Min difference (ps)
0.6	1.2	1039.8	647	843.4	196.4
0.5	1	911	502	706.5	204.5
0.42	0.84	775.6	427.1	601.35	174.25

These values and ranges may change with different other processes (100/70 nm process) and technology model. It is very important to calculate the minimum and maximum pulse width values for a particular process, finger width and

technology model. But it will not be a safe design if the pulse width is kept at minimum value. Circuit may be harmed with additional delay or timing variations for various reasons (even for NBTI). If boundary values (maximum and minimum) are used, it may face flipping failure even for an additional 1 ps timing variation. Considering all these predictive issues, we have introduced nominal pulse width, which is the middle value of maximum and minimum. The benefit of using nominal value is that even if circuit is facing unexpected timing variations, data flipping will not be harmed as we have ensured secure pulse width with valid region at both edges of the pulse. Even for +/- 50 ps of timing variations, flipping will not be harmed.

It is observed that for different W/L ratios, we have different security regions of flip pulse width. Here security region is known as the time difference between the nominal value and the boundary value. As time variations may appear, it is secure to design the cell with maximum security region to compete with maximum timing variation. Also, minimum nominal pulse width value will ensure fastest data flipping. Technology model and process should be chosen according to desired specifications of SRAM.

IV. Proposed SRAM Architecture

We have redefined the architecture of SRAM according to the proposed 7-T memory cell and data flipping technique. The architecture is illustrated in Fig. 4.

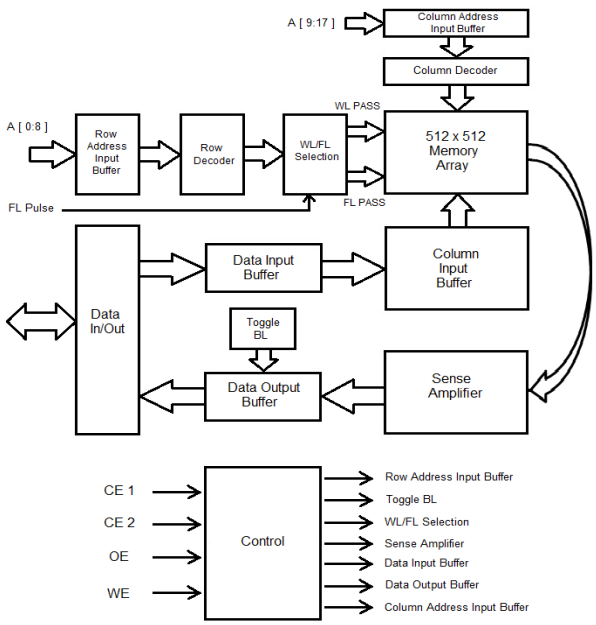


Fig. 4. Proposed SRAM Architecture

A. Read/Write/Standby/Flip Operation

Signal activity table of the proposed SRAM architecture is defined in Table II. This represents an active low SRAM system. The system is ready for read/write when both *CE1* and *CE2* are active low. During *CE1* and *CE2* are active low, depending on the logic in *WE* and *OE*, system performs read/write operations. For complimentary logic in *CE1* and *CE2*, device will be in standby mode or idle mode. In this mode, device can go on sleep/ idle state to save power (Fig. 5 (a)). An extra activity mode introduced is called *Flip Mode.* Only when *CE1* and *CE2* are both high logic, flip mode will be activated and data flipping will start (Fig. 5(b)).

TABLE II: SRAM Signal Activity Table (H = High Logic, L= Low Logic, X = Don't Care)

Mode	/CE1	/CE2	/WE	/OE	Data I/O
Stand-by	H	L	X	X	High Z
Stand-by	L	H	X	X	High Z
Flip	H	H	X	X	High Z
I/O Disable	L	L	H	H	High Z
Read	L	L	H	L	Data Out
Write	L	L	L	X	Data In

As active mode (read/write), standby mode and flip mode are activated in L-L, L-H/H-L and H-H pattern of *CE1* and *CE2* respectively, it will be logically secure for the system to allow one mode after another. For example, after active mode (L-L) is done, standby mode (L-H/H-L) should appear. Under no circumstances, flip mode (H-H) should appear just after active mode, as this may hamper data integrity. Here standby mode works as a secure boundary between active mode and flip mode. Also, it is possible to run the cycle *active-standby-flip-standby-active* by just changing one signal (*CE1* or *CE2)* as they follow gray code pattern (00-01-11-10-00).

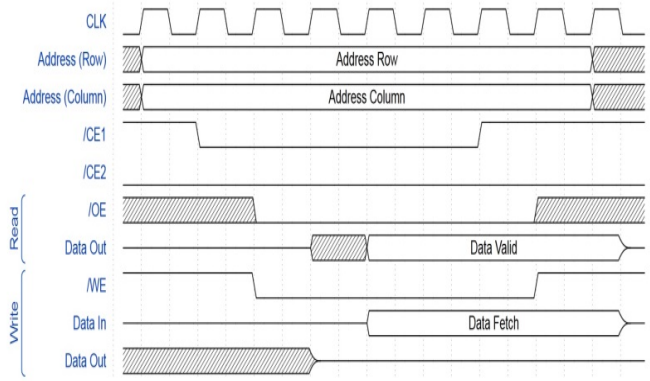


Fig. 5(a) SRAM Read/Write Operation

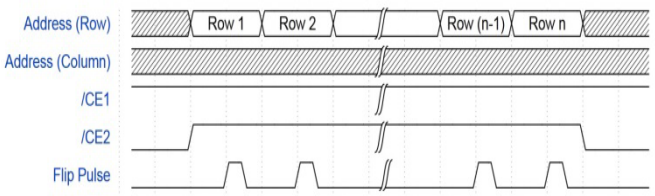


Fig. 5(b) SRAM Cell Flip Operation

B. Row Decoder as Flip Line Selector

The main issue in other data flipping techniques is that they require long time to flip all the cells. It is needed to individually access into every cell and flip it. Flip time increases with total number of cells in an array.

We have developed our proposed system in such a way that in one pulse, not only one cell will be flipped, rather all the cells in a row will be flipped. As in section IV, we have observed that our proposed 7-T cell just need a controlled pulse in the *flip MOS* to flip the data (regardless of what data is stored), it is possible to flip data of the entire row with just one flip pulse.

All the *flip transistors* in one row are connected with a single flip line. Each row has individual flip line. As active mode and flip mode are isolated from each other, row decoder can be used as row (WL) selector while in active mode and flip line (FL) selector while in flip mode. A mux based module (*WL/FL Selector)* should be implemented to select whether the row address will power the word line or flip line.

In this proposed technique, not each cell is flipped individually; rather all the cells in one row are flipped at the same time. During flip mode, processor generates incremental row address to flip the cells of one row after another. Row flip line is selected and flip pulse is provided to flip all the cells in that row. Total flip time depends on total number of rows, not total number of cells in the array. So, total flip time has been reduced in a significant manner. During active mode, row address will perform its conventional procedure (to select word line).

C. BL line select

After the data flipping is done, complimentary bit line will be selected to read/write the cell. A Module called Toggle BL consists of a toggle flip flop and Mux. When a flipping will occur, it will flip the active BL (if initially BL is selected to read/write data then after flipping BLB will be selected) to be active in active mode. The reason is to always get the data from the cell or write the data in the cell correctly. This will be in parallel with the flipping procedure so no extra cycle is needed for BL flipping.

D. Switching optimization

It is observed that it is possible to flip all the cells with one flip pulse if all the flip MOS in the array are connected with just one flip line. It will also save much time. As flipping is done inside the cell automatically so it will not harm the data of other cells during general flipping. But it is observed that if all the cells in an array are flipped at the same time, there will be a large switching current flow at one instant. Such large switching current can damage the whole chip. So, here it has been proposed to flip the cells of a particular row at a time. When flipping in one row is done, row decoder will select the next row to flip. After one flipping is done, system will get enough time to be electrically stable till next row flipping.

It is very important to take proper precautions as all the cells in one row are getting flipped at the same instant. Comparatively larger switching current will be drawn if all the cells in one row are flipped at the same time compared to flip the cells individually at individual instants. Switching current density for 180 nm process (1.8 V supply voltage) different flipping methodologies is described in Table III.

It is observed from Table III that switching density in proposed flipping methodology is quite larger than the conventional individual cell flipping methodology. It is necessary to implement extra ground ports/sink channels to maintain this extra switching current load at one instant. It is noticeable that total switching current is constant for all the methodologies. It is analyzed that how much switching current the system is dealing with at one flipping instant. For conventional method, switching current at one instant will be just the switching current of one cell. If switching current in one cell is 1 nA, then for our proposed methodology, total switching current at one instant will be 160 mA if total of 512 cells exist in one row, which is quite large but maintainable. But if all the cells are flipped at same instant, total switching current drawn at one instant for 512×512 array will be 82 A, which is a huge current load for present micro-system. General flipping of the whole array is avoided due to this alarming switching current density although general flipping is the fastest of all this methodologies. Considering both switching density and flipping time, proposed system is the most optimum choice for SRAM system design.

TABLE III: SWITCHING CURRENT DENSITY FOR DIFFERENT FLIPPING METHODOLOGIES (512×512 ARRAY)

Methodology	**Percentage of Total Switching Current at One Instant**	**Switching Current Load at One Instant**
Conventional Flipping	0.00038	313 μA
Proposed Flipping	0.1953	160 mA
Flipping Entire Cell at the Same Time	100	82 A

V. AREA VS. TIMING

In present world, device area is a very important factor. If one wants to propose a system with increased area, he should present strong evidence supporting his proposal. In case of memory circuit, area sensitivity is much more critical. As memory array is getting denser day by day, it is always needed to decrease the area as much as possible.

Although our SRAM design is significantly time friendly for data flipping purpose, some extra circuits, devices and blocks needed to be applied. An optimum condition has been reached in so that circuit area increase is acceptable and at the same time, system is faster than before. It is observed that although area has increased, but timing development is far

better than before that area increase can be considered to achieve such timing advantages.

A. Area

An extra NMOS has been introduced in each cell which increases the size of the memory cell, hence the memory array. This extra NMOS is needed to flip the data of the cells. So, the proposed 7-T SRAM cell is actually working as 6-T cell at normal operation. At flip mode, the extra NMOS (flip MOS) is needed to flip the data stored in the cell. For a 512×512 memory array, total array size has increased 16.67 % according to our proposed cell design.

For this proposed data flipping technique, some extra block has been introduced in this system. *WL/FL Selector* has to be included which will actually select WL in normal operation and FL at flip operation. Also to flip the active bit line after each successful flipping, we have introduced another block which will toggle the bit line. This block consists of only a toggle flip flop which will change its stored value at every flip pulse so that complementary bit line is selected at after each flip operation. It seems that the increase in the peripheral circuits of the system is not so alarming. In fact the hardware approach described in [3] consists more peripheral circuits than the proposed architecture in Fig. 4. Although, it has been ensured that less increase in peripheral circuits, the present memory array covers more area. It has been discussed in the next point that in spite of increase in memory array, the present system provides satisfactory timing development.

B. Timing

Flipping approach described in [3] requires comparatively long time to flip the data of the whole array. System needs to access in each cell individually, flip the data with an external inverter and put that data in the cell again. This procedure is a very time consuming for present gigahertz systems. It seems that for each data flipping according to described hardware approach in [3], if read and write cycle times are each considered 50 ns, approximately 100 ns will be needed to flip the data of one cell of the array. It is because the conventional flipping techniques are programmed to read the data of a cell (read cycle), invert it and write the inverted data back to the cell (write cycle).

TABLE IV: TOTAL FLIP TIME FOR DIFFERENT MEMORY ARRAY SIZE

Flipping Technique	**Flipping time (one cell) (ns)**	**Total Flipping Time in ms (Different Memory Size, kB)**		
		8	**32**	**131**
Conventional	100	6.55	26.2	104.86
Proposed	25	1.64	6.55	26.21

The proposed system does not select individual cells to flip rather it selects an entire row and flip all the cells in that row. In conventional flipping techniques, flipping time depends on the cell size. If cell size increases, flip time increases. In the proposed system, flip time depends on total number of rows. Flip time increases with the increasing number of rows in the array. So increase of flipping time with increase in cell size is also very low compared to conventional system. As system (systems with read/write cycle time of 50 ns) needs to access to the row only, it will need only 25 ns to flip the data of all cells in entire row (columns are not selected and cell is not accessed for read and write operation).

Table IV shows the increase of flip time with increase in cell array size for both conventional and proposed flipping technique (read/write cycle time 50 ns).

It is observed that not only the flipping time has been reduced, but also increase in flipping time is reduced with increasing cell size. This significant timing development for flipping the array ensured less possibility of system interrupt on normal operation for flipping procedure. As total flipping time is much smaller, it makes the flipping mode very time friendly for the system and ensure working on multiple flipping times.

VI. CONCLUSION

A faster cell data flipping approach has been introduced here to mitigate the degradation of circuit parameters in SRAM. Although the proposed data flipping technique increases overall cell array and peripheral area, it provides very low flipping time for the entire array. As flipping mode appears interrupting the normal operation mode, longer flipping time is not convenient for system with high activity factor. So, the method proposed in this paper offers a faster approach to periodic data flipping of SRAM array for NBTI recovery. Data flipping technique is a very effective process to mitigate NBTI degradation as it provides recovery period after each stress period. Due to long flipping time in conventional design, this technique is avoided in high activity and high frequency systems. As the proposed system ensures very low flipping time, it will be convenient to utilize proposed data flipping technique for NBTI recovery even for fast systems with high activity.

The 7-T memory cell in this work is not like well-known 7-T cell which is designed for better SNM performance. An extra *flip MOS* has been used which is active while in flip mode and remains inactive while in normal operation mode. So, the SNM performance of this proposed cell is not comparable to that of a conventional 7-T cell. Also, the proposed data flipping process introduces a comparatively large switching current density.

With the modern processor technology, total flipping time may be developed using multiple processor system, even if individual cell flipping is applied. But practically it is always beneficial to utilize the processor as best as possible. If the proposed approach can be applied even at a multiple processor system, total flipping time will be much lower than individual cell flipping systems. Processors can perform other necessary multi-tasking operation and provide much better timing performance as the proposed system can save a significant amount of time. It is already described in section II about the individual cell flipping approaches which may take lots of

processor cycles. This delay can definitely be minimized by applying the proposed approach. Although having area constraint, the proposed system will provide much better timing.

REFERENCES

[1] D. K. Schroder, "Negative bias temperature instability: what do we understand?," *Microelectronics Reliability,* vol. 47, no. 6, pp. 841-852, Jun. 2007.

[2] S. V. Kumar, C. H. Kim, and S. S. Sapatnekar, "Impact of NBTI on SRAM read stability and design for reliability," in *Proc. IEEE 7th Int. Sym. on Quality Electronic Design,* Mar. 2006, pp. 218.

[3] E. H. Poindexter and P. J. Caplan, "Characterization of Si/SiO_2 interface defects by electron spin resonance," *Progress in Surface Science*, vol. 14, no. 3, pp. 201-294, 1983.

[4] Y. Miura and Y. Matukura, "Investigation of silicon-silicon dioxide interface using MOS structure," *Japanese Journal of Applied Physics,* vol. 5, no. 2, pp. 180-180, 1966.

[5] D. K. Schroder and J. A. Babcock, "Negative bias temperature instability: Road to cross in deep submicron silicon semiconductor manufacturing," *Journal of Applied Physics*, vol. 94, no. 1, pp. 1-18, Jul. 2003.

[6] V. Reddy, et al. "Impact of negative bias temperature instability on digital circuit reliability," *Microelectronics Reliability,* vol. 45, no. 1, pp. 31-38, Jan. 2005.

[7] S. Mahapatra, P. B. Kumar, and M. A. Alam, "Investigation and modeling of interface and bulk trap generation during negative bias temperature instability of p-MOSFETs," *IEEE Transactions on Electron Devices*, vol. 51, no. 9, pp. 1371-1379, Sep. 2004.

[8] A. T. Krishnan, et al. "NBTI impact on transistor and circuit: models, mechanisms and scaling effects [MOSFETs]," in *Proc. IEEE Int. Electron Devices Meeting,* Dec. 2003, pp. 14.5.1-14.5.4.

[9] V. Reddy, J. Carulli, A. Krishnan, W. Bosch, and B. Burgess, "Impact of negative bias temperature instability on product parametric drift," in *Proc. Int. Test Conference,* Oct. 2004, pp. 148-155.

[10] A. S. Goda, and G. Kapila, "Design for degradation: CAD tools for managing transistor degradation mechanisms," in *Proc. IEEE 6th Int. Sym. on Quality Electronic Design,* Mar. 2005, pp. 416-420.

[11] B. C. Paul, K. Kunhyuk, H. Kufluoglu, M.A. Alam, and K. Roy, "Impact of NBTI on the temporal performance degradation of digital circuits," *IEEE Electron Device Letters*, vol. 26, no. 8, pp. 560-562, Aug. 2005.

[12] X. Yang, E. Weglarz, and K. Saluja, "On NBTI degradation process in digital logic circuits," in *Proc. IEEE 20th Int. Conf. on VLSI Design,* Held jointly with *6th Int. Conf. on Embedded Systems,* Jan. 2007, pp. 723-730.

[13] C-H. Hua, T-S. Cheng, and W. Hwang, "Distributed data-retention power gating techniques for column and row co-controlled embedded SRAM," in *Proc. IEEE Int. Work. on Memory Technology, Design, and Testing,* Aug. 2005, pp. 129-134.

[14] T. Siddiqua, and S. Gurumurthi, "Enhancing NBTI recovery in SRAM arrays through recovery boosting," *IEEE Transactions on VLSI Systems*, vol. 20, no. 4, pp. 616-629, Apr. 2012.

[15] B. Zhu, J. S. Suehle, Y. Chen, and J. B. Bernstein, "Negative bias temperature instability of deep sub-micron p-MOSFETs under pulsed bias stress," in *Proc. IEEE Int. Work. on Integrated Reliability,* Oct. 2002, pp. 125-129.

[16] B. Zhu, J. S. Suehle, J. B. Bernstein, and Y. Chen, "Mechanism of dynamic NBTI of pMOSFETs," in *Proc. IEEE Int. Work. on Integrated Reliability,* Oct. 2004, pp. 113-117.

[17] M. A. Alam, "A critical examination of the mechanics of dynamic NBTI for PMOSFETs," in *Proc. IEEE Int. Electron Devices Meeting,* Dec. 2003, pp. 14.4.1-14.4.4.

[18] S. Mahapatra, P. B. Kumar, T. R. Dalei, D. Sana, and M. A. Alam, "Mechanism of negative bias temperature instability in CMOS devices: degradation, recovery and impact of nitrogen," in *Proc. IEEE Int. Electron Devices Meeting,* Dec. 2004, pp. 105-108.

[19] C. Schlünder, H. Reisinger, W. Gustin, and T. Grasser, "A new physics-based NBTI model for DC and AC stress enabling accurate circuit aging simulations considering recovery," in *Proc. Zuverlässigkeit und Entwurf,* Sep. 2010, pp. 33-40.

[20] M. A. Alam and S. Mahapatra, "A comprehensive model of PMOS NBTI degradation," *Microelectronics Reliability*, vol. 45, pp. 71-81, 2005.

[21] M. A. Alam, H. Kufluoglua, D. Varghesea and S. Mahapatra, "A comprehensive model for PMOS NBTI degradation: Recent progress," *Microelectronics Reliability*, vol. 47, no. 6, pp. 853-862, Jun. 2007.

[22] G. Chen, et al., "Dynamic NBTI of PMOS transistors and its impact on device lifetime," in *Proc. IEEE 41st Annual Int. Sym. on Reliability Physics,* Mar. 2003, pp. 196-202.

Face Recognition using Eyes, Nostrils and Mouth Features

Sushil Kumar Paul, Mohammad Shorif Uddin

Department of Computer Science and Engineering
Jahangirnagar University, Dhaka-1342, Bangladesh
E-Mail: paulksushil@yahoo.com, shorifuddin@gmail.com

Saida Bouakaz

LIRIS Lab, University Claude Bernard Lyon1,
69622 Villeurbanne Cedex, France
E-Mail: saida.bouakaz@liris.cnrs.fr

Abstract—**This paper describes a face recognition algorithm that extracts the eyes, nostrils and mouth features from cumulative distribution function (CDF) by applying Otsu thresholding. The algorithm, which is inspired by the probability of white pixels of binary facial image, has been tested using the BioID frontal face large database in different illuminations, expressions and lighting conditions. Illumination and lighting variations are addressed using a selective equalization technique. The experimental results have confirmed an average recognition rate of 93.55%.**

***Keywords*-Cumulative distribution function, Otsu thresolding, probability of white pixels, facial features, face recognition.**

I. INTRODUCTION

Face recognition is one of the most flourishing areas in computer vision like identification, authentication, security, surveillance system, human-computer interaction, psychology and so on [1].Eyes, nostrils, and mouth are the key features for face recognition [2], [3]. Eyes and mouth bear some form of facial expressions and eyes are the most crucial facial feature because of its inter-ocular distance, which is constant among people and unaffected by moustache or beard [3]. Another valuable face feature is nostrils because nose tip has the symmetry point between right- and left-side face regions and it indicates the head pose and is not affected by facial expressions [4]. Therefore, face recognition is distinctly influenced by these features.

Extensive research works have been performed in this domain. Currently, active-shape model (ASM) and active-appearance model (AAM) are widely used techniques for face recognition and tracking [5]. Facial feature extraction is a prerequisite for face recognition. There are two categories of facial feature extraction strategies: texture-based and shape-based. Texture-based methods take local texture e.g. pixel values around a given specific feature point instead of concerning all facial feature points as a shape. Some texture-based facial feature extraction algorithms are: hierarchical 2-level wavelet network [6], log Gabor wavelet network [7], neural-network-based eye-feature detector [8]. Shape-based facial feature extraction algorithms includes active wavelet network [9], direct appearance model [10], component-based with 3D morphable models [11] etc. In addition, some hybrid techniques are also developed, such as elastic bunch graph matching [12], AdaBoost with shape constraints [13], 3D probabilistic approach [14] etc.

In the above mentioned reference [14] Wiskott et al. represent faces by a rectangular graph which is based on Gabor wavelet transform and each node is labelled with a set of complex Gabor wavelet coefficients, in reference [13] Cristinacce and Cootes used the Haar wavelet based AdaBoost classifier combined with the statistical shape model and in reference [11] B. Weyrauch at al. applied the 14 facial components, which is based on 3D geometrical model. In both ASM and AAM, a model is built for predefined points by using the test images and then an iterative scheme is applied to this model in detecting feature points. Most of the above mentioned algorithms are not entirely reliable due to variation in pose, illumination, facial expression, and lighting condition and high computational complexity. So, it is indispensable to develop a robust, automatic, and accurate facial-feature extraction as well as face recognition algorithm, which is capable of coping above mentioned diverse situations. This paper tries to fill this gap. Here, we propose a robust face recognition algorithm that extracts the eyes, nostrils and mouth features from cumulative distribution function (CDF) by applying Otsu's optimum global threshold value [15] instead of multiple threshold values [16],[17],[18],[19].

The remainder of this paper is organized as follows. Section II describes the proposed technique. Section III shows the experimental results and finally conclusions are drawn in Section IV.

II. PROPOSED METHODOLOGY

A. Feature Extraction

Figure 1 shows the simplified flow diagram of our proposed algorithm that includes preprocessing, main processing and recognition blocks. The preprocessing block detects and crops the face region, normalizes the face size, removes the forehead region from then normalized face and then equalizes the face image without forehead area (WFA) if the average pixel intensity value is less than 170 [20], [21]. This is done to cope with the illumination and lighting variations in a selective way. The processing block is responsible for converting the greyscale face image WFA to a binary image using Otsu thresholding on the basis of CDF [3], [15]. Then it extracts four ROIs (region of interest), such as right eye, left eye, nose, and mouth areas for both gray level and binary images (see Fig. 2). The details are explained in reference [18], [19].

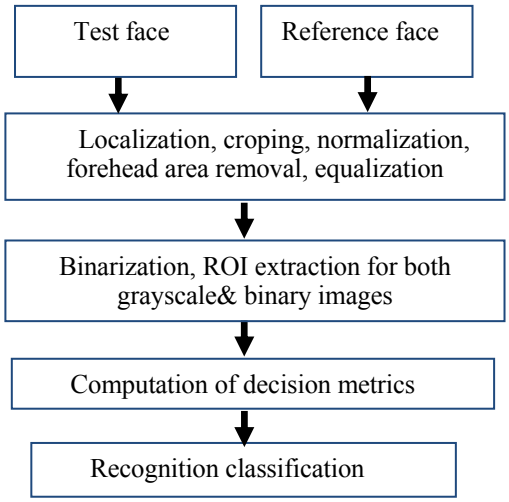


Figure 1. Simplified block diagram of the proposed face recognition algorithm.

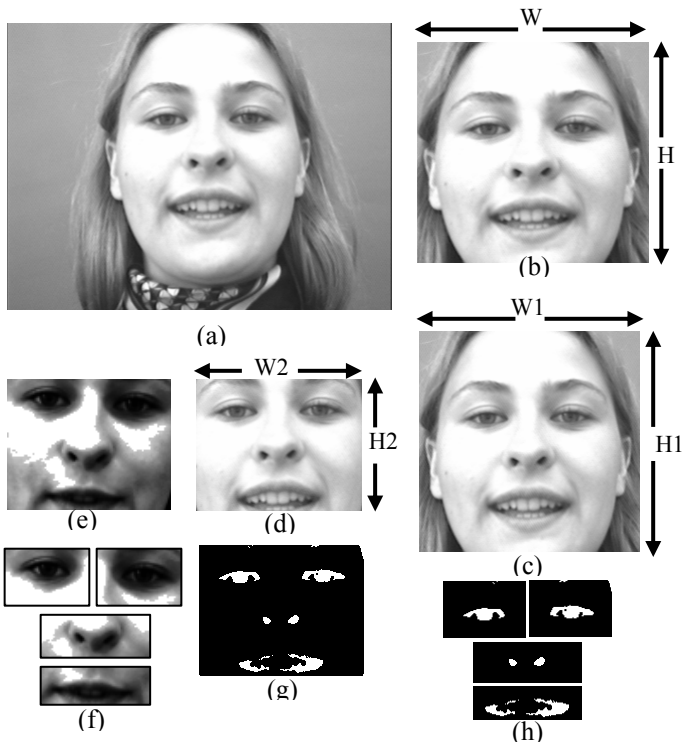


Figure 2. Processing steps: (a) input image, (b) detected and cropped the face with forehead, (c) normalized face with forehead region (Size: W1×H1=128×128 pixels), (d) face without forehead (size=W2×H2=96×76 pixels), (e) histogram equalized image, (f) gray level face image is divided into three vertical parts, which are eyes, nose and mouth areas , (g) conversion of image (d) or (e) to binary, (h) binary face image is divided into three vertical parts, which are eyes, nose and mouth areas.

Figure 2 presents the outputs of the different steps for a typical face image. Next subsection describes the recognition algorithm.

$$CDF_{I_{gray}(x,y)}(v) = \sum_{j=0}^{v} P_{I_{gray}(x,y)}(j) \tag{1}$$

$$I_{bin}(x,y) = \begin{cases} 255 & \text{if } CDF_{I_{gray}(x,y)} \le T_{Otsu} \\ 0 & otherwise \end{cases} \tag{2}$$

Where,

$I_{gray}(x,y)$ is the gray level intensity of face image WFA and

$P_{Igray(x,y)}(v)$ is the distribution function, i.e. the probability of an occurrence of a pixel having intensity level v.

B. Recognition Approach

To reach a recognition decision for a test face, we computed the values of the following three metrics for five images (equalized WFA face image and the four ROIs i.e. five components). These are:

1. Standard deviation of Chi-square value (SD_{Chi}) between the binary form of test image and a reference database image. It is given in Eq. (3).

$$SD_{Chi}(I_{ref}(x,y), I_{test}(x,y)) = \frac{\sqrt{\sum_{i=1}^{5}(X_i - X_{avg})^2}}{5} \tag{3}$$

$$\begin{aligned} X_i &= Chi(I_{ref}(x,y), I_{test}(x,y)) = \\ &\frac{(P_{ref}(I_{ref}(x,y)) - P_{test}(I_{test}(x,y)))^2}{(P_{ref}(I_{ref}(x,y)) + P_{test}(I_{test}(x,y)))} + \\ &\frac{(Q_{ref}(I_{ref}(x,y)) - Q_{test}(I_{test}(x,y)))^2}{(Q_{ref}(I_{ref}(x,y)) + Q_{test}(I_{test}(x,y)))} \\ &= \frac{(P_{ref}(I_{ref}(x,y)) - P_{test}(I_{test}(x,y)))^2}{(P_{ref}(I_{ref}(x,y)) + P_{test}(I_{test}(x,y)))} + \\ &\frac{(P_{ref}(I_{ref}(x,y)) - P_{test}(I_{test}(x,y)))^2}{(2 - P_{ref}(I_{ref}(x,y)) - P_{test}(I_{test}(x,y)))} \end{aligned} \tag{4}$$

Where,

$P = \frac{No\ of\ white\ pixels}{Image\ size}$ $P = \frac{Noofwhitepixels}{Imagesize}$, and Q=1-P. $P_{ref}(I_{ref}(x,y))$ and $P_{test}(I_{test}(x,y))$ are the probability of white pixels of reference and test images, respectively. X_{avg} is the average of X_i's (i = 1, 2, 3, 4, 5).

2. Absolute difference of probability of white pixels (ADPWP) between the binary form of test image and a reference database image. It is shown in Eq. (5).

$$ADPWP_i(I_{ref}(x,y), I_{test}(x,y)) = \left| P_{ref}(I_{ref}(x,y)) - P_{test}(I_{test}(x,y)) \right| \quad (5)$$

3. Euclidean distance metric (EDM) between the grayscale form of test image and a reference database image. It is expressed in Eq. (6).

$$EMD_i(I_{ref}(x,y), I_{test}(x,y)) = \frac{\sqrt{\sum_{x,y} (I_{ref}(x,y) - I_{test}(x,y))^2}}{W \times H} \quad (6)$$

Where, W and H represent the image region's width and height, respectively.

If the value of SD_{Chi} is less than the threshold value (T_{Recog}) AND any one out of five values of $ADPWP_i$ is less than or equal to 0.005 AND any one out of five values of EDM_i is less than 1.0, then the reference and the test images are the same image.

III. EXPRIMENTATIONS

A. Face Database

In this paper all experimental works are done using head-shoulder BioID face large database [22]. The dataset with the frontal view of a face of one out of 25 different test persons consists of 1521 gray level images having properties of different illumination, face area, complex background with a resolution of 384×286 pixels.

B. Results

The proposed algorithm has been developed and tested on Code::Blocks the open source, cross-platform combine with c/c++ programming language, and GNU GCC compiler. Some OpenCV library functions were used for face detection and localization, cropping, and image normalization purpose [20], [22].The threshold (T_{Recog}) for SD_{Chi} is taken as 0.008 for BioID database.

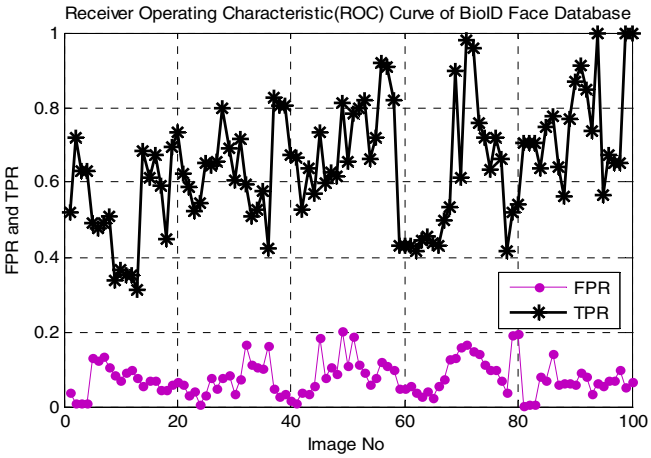


Figure 3: ROC curve for 100 images (considered all persons of entire BioID database)

Figure 3 is shown the Receiver Operating Characteristic (ROC) curves and figure 4 is shown the recognition accuracy for 100 images (Considered entire database, the maximum 8 images/person and minimum 1 image/person) are taken for TPR(True Positive Rate)≥0.50 except all images of person no 4 and 11) for BioID database. The values of TPR for most of persons are shown the near or above 0.50 and the False Positive Rate (FPR) is always below the TPR curve. Table I indicates the results of our face recognition algorithm, where the overall average recognition rate is 93.55%. We compared our algorithm with B. Weyrauch, et al. [11].The comparison results are shown in table II.

Some of the recognition results are shown in the Figure 4. Table II shows an average recognition accuracy, which is better than the B. Weyrauch, et al. [11].

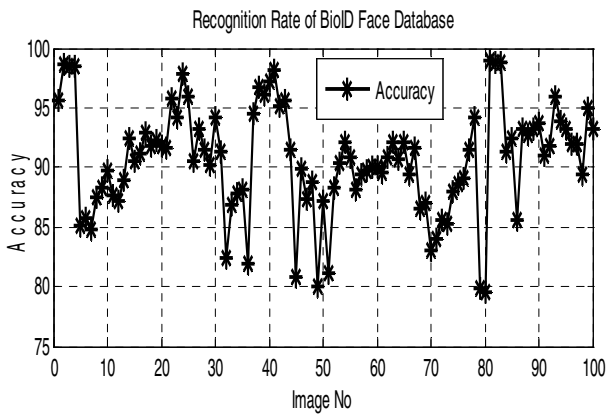


Figure 4: Recognition Rate for 100 images (Considered all persons of entire BioID database)

TABLE I. RECOGNITION ACCURACY

Face Database	No of Components are Used	Threshold Value For CDF	Threshold Values For Face Recognition	Best Recognition Accuracy and Corresponding TPR		Average Recognition Accuracy
				Accuracy	TPR	
BioID [21]	5	$T_{otsu}/6.0$	$SD_{Chi}<T_{\text{Recog}}$ $T_{\text{Recog}}=0.008$ $ADPWP_i \le 0.005$ $EDM_i<1.0$	98.93 %	0.7059	**93.55 %**

TABLE II. COMPARISON OF OUR RESULT WITH COMPONENT BASED ALGORITHM[11]

Methods	No of Components are Used	Recognition Accuracy
Component Based	14	**88.00%**
Ours	5	**93.55%**

IV. Conclusions

In this paper, we have shown how salient facial features, such as eyes, nostrils and mouth are extracted from the cumulative distribution function (CDF) by applying Otsu's optimum global thresholding technique. We then linked these features to a face recognizer by computing three metrics. We have overcome the illumination and lighting variations in a selective way. Finally we tested our algorithm using the large gray level BioID face database having properties of different illumination, facial expression, and lighting condition. We achieved an average recognition accuracy of 93.55%. We will focus our future concentration on improving accuracy as well as TPR by accommodating weighted feature values.

References

[1] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face Recognition: A Literature Survey," ACM Computing Surveys, Vol. 35, No. 4, December 2003.

[2] D. Vukadinovic and M. Pantic," Fully Automatic Facial Feature Point Detection Using Gabor Feature Based Boosted Classifiers,"IEEE nternational Conference on Systems, Man and Cybernetics, Hawaii, October 10-12, 2005.

[3] Mansour Asadifard, and Jamshid Shanbezadeh, "Automatic Adaptive Center of Pupil Detection Using Face Detection and CDF Analysis," Proceedings of The International Multi Conference of Engineers and Computer Scientists 2010 , pp130-133 , IMECS 2010, Hong Kong, 17-19 March, 2010.

[4] Wei Jen Chew, Kah Phooi Seng, and Li-Minn Ang, "Nose Tip Detection on a Three-Dimensional Face Range Image Invariant to Head Pose,"Proceedings of The International Multi Conference of Engineers and Computer Scientists 2009 , Vol I , IMECS 2009, Hong Kong, March 18-20, 2009.

[5] I. Matthews and S. Baker, "Active Appearance Models Revisited," Int'l Journal Computer Vision, vol. 60, no. 2, pp.135-164, 2004.

[6] R.S. Feris, et al., "Hierarchical Wavelet Networks for Facial Feature Localization," Proc. IEEE Int'l Conf. Face and Gesture Recognition, pp. 118-123, 2002.

[7] E. Holden, R. Owens, "Automatic Facial Point Detection," Proc. The 5th Asian Conf. on Computer Vision, 23-25 January 2002, Melbourne, Australia.

[8] M. J. T. Reinders, et al., "Locating Facial Features in Image Sequences using Neural Networks," Proc. IEEE Int'l Conf. Face and Gesture Recognition, pp230-235, 1996.

[9] C. Hu, et al.,"Real-time view-based face alignment using active wavelet networks," Proc. IEEE Int'l Workshop Analysis and Modeling of Faces and Gestures, pp. 215-221, 2003.

[10] S. Yan, et al., "Face Alignment using View-Based Direct Appearance Models," Int'l J. Imaging Systems and Technology, vol. 13, no. 1, pp. 106-112, 2003.

[11] B. Weyrauch, et al, "Component-based Face Recognition with 3D Morphable Models," 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'04), June 27- July 2, 2004, Washington DC, USA.

[12] L. Wiskott, et al., "Face Recognition by Elastic Bunch Graph. Matching," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 19, no.7, pp. 775-779, 1979.

[13] D. Cristinacce, T. Cootes, "Facial Feature Detection Using AdaBoost with Shape Constrains", British Machine Vision Conference, 2003.

[14] L. Chen, et al., "3D Shape Constraint for Facial Feature Localization using Probabilistic-like Output", Proc. IEEE Int'l Workshop Analysis and Modeling of Faces and Gestures, pp. 302-307, 2004.

[15] Nobuyuki Otsu, "A threshold selection method from gray-level histograms,"IEEE Trans. Sys., Man., Cyber. vol. **9** (1), pp. 62–66, (1979).

[16] SushilKumarPaul, Saida Bouakaz and Mohammad Shorif Uddin, "Automatic Adaptive Facial Feature Extraction using CDF Analysis", Proceedings of The International Conference on Digital Information and Communication Technology and its Applications, Communications in Computer and Information Science, Volume 166, Part 1, Pages 327-338, DICTAP2011, 21-23 June 2011, Dijon, France.

[17] Sushil Kumar Paul, Mohammad Shorif Uddin and Saida Bouakaz "Extraction of Facial Feature Points using Cumulative Histogram,"IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 1, No. 3, pp. 44-51, January 2012, http://www.IJCSI.org

[18] Sushil Kumar Paul, Mohammad Shorif Uddin and Saida Bouakaz, "Extraction of Facial Feature Points Using Cumulative Distribution Function by Varying Single Threshold Group," Proceedings of the IEEE/OSA/IAPR International Conference on Informatics, Electronics and Vision (ICIEV2012),18-19 May, 2012, Dhaka, Bangladesh, pp. 806-811.

[19] Sushil Kumar Paul, Mohammad ShorifUddin and Saida Bouakaz, "Extraction of Facial Feature Points from CDF using Otsu Thresholding,"Journal of Imaging and Vision Research, vol. 1, no. 2, pp. 7-13, December 2012,http://www.casr-journals.org

[20] P. Viola and M. J. Jones, "Robust Real-time Object Detection", *International Journal of Computer Vision*, Vol. 57, No.2, p.137-154, 2004.

[21] http://sourceforge.net/projects/opencvlibrary/files/opencvwin/2.0/OpenCV-2.0.0a-win64.exe/download

[22] BioID Face Database, Avialable online: http://www.bioid.com/downloads/facedb/index.php

Performance Comparison of Constant Coefficient and Variable Coefficient Tracking Filters in Radar Detection

Gp Capt Dr Mohammed Hossam-E-Haider [(1)], Asif Al Faisal[(2)], Nazmul Rahim[(3)] and Md. Saiduzzaman[(4)]
[(1-4)]Electrical, Electronic and Communication Engineering Department
Military Institute of Science and Technology (MIST)
Dhaka, Bangladesh.
[(1)]haider8400@yahoo.com, [(2)]faisal_mist09@yahoo.com, [(3)]nazmuleece7@yahoo.com, [(4)]sajibeece09@yahoo.com

***Abstract*—In this paper comparison between constant coefficient and variable coefficient tracking filters for radar has been presented. αβγ filter and Kalman filter are discussed as instances of constant coefficient and variable coefficient filters respectively. Both of the filters are sophisticated smoothing and prediction tracking filters which are used in sampled data target trackers. As polynomial predictor linear recursive filters they can construct future position based on present position measurements. For the same target model the mean residual error from both of the filters are presented to compare their relative performance. Noise variance was varied to observe its effect on both filters. Performance of αβγ filter becomes deteriorated with increased noise whereas Kalman filter can track a target with high accuracy even with increased noise variance. These clearly illustrate the advantage of using a variable coefficient Kalman filter over a fixed coefficient αβγ filter.**

***Index Terms*—Radar tracker, Variable coefficient, Constant coefficient, Alpha-beta-gamma (αβγ) filter, Transition matrix (Φ), Kalman filter.**

I. INTRODUCTION

Radar provides information including the bearing and range (and therefore position) of the object from the radar scanner. The information can be used in many different fields, a major one of which is in target tracking. Target tracking is the process of locating a moving object throughout a sequence of consecutive frames. In military radars, tracking is a must for fire control and missile guidance. Commercial radar systems, such as civilian airport traffic control radars, may utilize tracking as a means of controlling incoming and departing airplanes. Due to its ability of holding many targets in track simultaneously, multiple target radar tracker has a vast range of practical applications.

Radar is an object-detection system which uses radio waves to determine the range, altitude, direction, or speed of objects. It can be used to detect aircraft, ships, spacecraft, guided missiles, motor vehicles, weather information and terrain. The radar antenna transmits pulses of radio waves which bounce off any object in their path. The object returns a tiny part of the wave's energy to the receiver. By observing the time for a round trip of the pulse, the position (or distance) of the target is calculated. Position of the target is observed for consecutive sample times to predict the next position of the target, i.e., target tracking.

Multiple target trackers or track-while-scan radars use sophisticated smoothing and prediction filters to estimate the target parameters between scans. This prediction filters can be constant coefficient or variable coefficient. The most common example of constant coefficient filter is the αβγ filter and its variations. Kalman filter has variable gain coefficients which are computed dynamically on the basis of minimizing the mean squared error. It can predict future estimations even when the precise nature of the modeled system is unknown.

In this paper, the superiority of Kalman filter over αβγ filter in radar tracking has been shown. Radar performance is measured on the basis of minimizing mean residual error. Fundamental estimation filters and performance metrics are simulated using MATLAB to provide relevant plots for study.

II. RADAR TRACKER

A radar tracker is a component of a radar system, or an associated command and control system, that associates consecutive radar observations of the same target into tracks. It is particularly useful when the radar system is reporting data from several different targets or when it is necessary to combine the data from several different radars or other sensors. With the aid of sophisticated computer systems, multi-function (detection, tracking and discrimination) radars are capable of simultaneously tracking many targets. Each target is sampled once (mainly range and angular position) during a dwell interval (scan). Then, by using smoothing and prediction techniques future samples can be estimated. Radar systems that can perform multi-tasking and multi-target tracking are known as Track-While-Scan (TWS) radars.

Once a particular target is detected, the radar may transmit up to a few pulses to verify the target parameters, before it establishes a track file for that target. Target position, velocity, and acceleration comprise the major components of the data maintained by a track file. Once a Track-While-Scan (TWS)

radar detects a new target it initiates a separate track file for that detection; this ensures that sequential detections from that target are processed together to estimate the target's future parameters. Position, velocity and acceleration comprise the main components of the track file. Typically, at least one other confirmation detection (verify detection) is required before the track file is established. The TWS system places a gate around the target position and attempts to track the signal within this gate. The gate dimensions are normally azimuth, elevation, and range. More precisely, targets must stay within the gate boundary during successive scans.

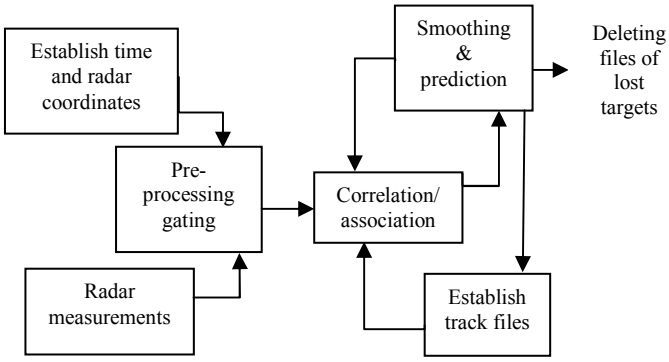


Fig. 1. Simplified block diagram of TWS data processing.

There are many different mathematical algorithms used for implementing a radar tracker of varying levels of sophistication. However, they all perform steps similar to the following every time the radar updates:

- Associate a radar plot with an existing track (plot to track association).
- Update the track with this latest plot (track smoothing).
- Spawn new tracks with any plots that are not associated with existing tracks (track initiation).
- Delete any tracks that have not been updated, or predict their new location based on the previous heading and speed (track maintenance).

Track-while-scan radar systems sample each target once per scan interval, and use sophisticated smoothing and prediction filters to estimate the target parameters between scans. To this end, the constant coefficient αβγ filter and variable coefficient Kalman filter are commonly used.

III. Constant Coefficient $\alpha\beta\gamma$ Filter

αβγ filter is a one-dimensional third order, fixed gain, polynomial predictor/corrector linear recursive filter. In simultaneously tracking many target each target is sampled once (mainly range and angular position) during a dwell interval (scan). Then by using αβγ filter future samples can be estimated. This filter can reconstruct position, velocity and constant acceleration based on position measurements.

For αβγ filter state vector representing range is given by

$$\underline{x} = \begin{bmatrix} R \\ \dot{R} \\ \ddot{R} \end{bmatrix} \tag{1}$$

Where $R, \dot{R}$ and $\ddot{R}$ are respectively the range measurement, range rate (velocity) and acceleration. In αβγ filter the state transition matrix assists predicting the next state. The ideal transition matrix elements can be formulated as

$$\underline{\phi}[ij] = \begin{Bmatrix} T^{j-i} \div (j-i)! & 1 \le i, j \le n \\ 0 & j < i \end{Bmatrix} \tag{2}$$

The state transition matrix for the αβγ filter is

$$\underline{\phi} = \begin{bmatrix} 1 & T & \frac{T^2}{2} \\ 0 & 1 & T \\ 0 & 0 & 1 \end{bmatrix} \tag{3}$$

Using the transition matrix the next state is predicted as

$$\underline{x}(n+1 \mid n) = \underline{\phi}\,\underline{x}(n \mid n) \tag{4}$$

By inspection the αβγ filter has Gain matrix

$$\underline{K} = \begin{bmatrix} \alpha \\ \beta / T \\ \gamma / T^2 \end{bmatrix} \tag{5}$$

Zero mean Gaussian noise matrix

$$\underline{G} = [1\ 0\ 0] \tag{6}$$

αβγ filter produces the predicted position and velocity for the (n+1)th observation. It follows an input whose acceleration is constant with zero steady state error. In order to reduce the error at the output of the tracker, a weighted difference between the measured and predicted values is used in estimating the smoothed position, velocity and acceleration.

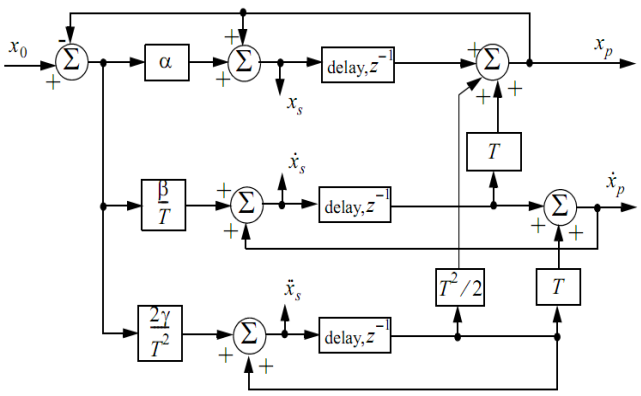


Fig. 2. An implementation for an αβγ tracker.

The equations for estimating the smoothed position, velocity and acceleration is [1]

$$x_s(n) = x_p(n) + \alpha(x_0(n) - x_p(n)) \tag{7a}$$

$$\dot{x}_s(n) = \dot{x}_s(n-1) + T\ddot{x}_s(n-1) + \frac{\beta}{T}(x_0(n) - x_p(n)) \tag{7b}$$

$$\ddot{x}_s(n) = \ddot{x}_s(n-1) + \frac{2\gamma}{T^2}(x_0(n) - x_p(n)) \tag{7c}$$

$$x_p(n+1) = x_s(n) + T\dot{x}_s(n) + \frac{T^2}{2}\ddot{x}_s(n) \tag{7d}$$

An important sub-class of the αβγ tracker is the critically damped filter, often called the fading memory filter. In this case coefficients are chosen on the basis of a smoothing factor (ξ), where $0 \le \xi \le 1$. The gain coefficients are given by [1]

$$\alpha = 1 - \xi^3 \tag{8a}$$

$$\beta = 1.5(1-\xi^2)(1-\xi) = 1.5(1-\xi)^2(1+\xi) \tag{8b}$$

$$\gamma = (1-\xi)^3 \tag{8c}$$

Where heavy smoothing takes place when $\xi \to 1$ and no smoothing means $\xi \to 0$.

For tracking systems with stationary measurement noise, maneuvering targets with constant acceleration can be accurately tracked with an αβγ filter. However, the accelerations of a maneuvering target are seldom constant in the tracking frame during the maneuver.

IV. Variable Coefficient Kalman Filter

The Kalman filter is a linear estimator that minimizes the mean squared error as long as the target dynamics are modelled accurately. All other recursive filters, such as the αβγ and the Benedict-Bordner filters, are special cases of the general solution provided by the Kalman filter for the mean squared estimation problem. It maintains real-time estimates of a number of parameters of a system, such as its position and velocity that may continually change. The estimates are updated using a stream of measurements that are subject to noise.

A key assumption in the mathematics of the Kalman filter is that measurement equations (i.e. the relationship between the radar measurements and the target state) and the state equations (i.e. the equations for predicting a future state based on the current state) are linear.

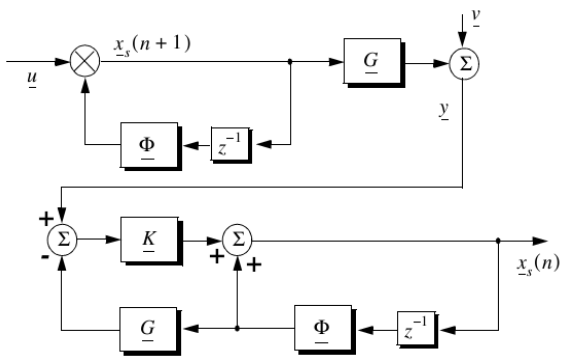


Fig. 3. Structure of the Kalman filter.

Fig. 3 shows a block diagram for the Kalman filter. The filtering equation is [1]

$$\underline{x}(n \mid n) = \underline{x}_s(n) = \underline{x}(n \mid n-1) + K(n)[y(n) - \underline{G}\underline{x}(n \mid n-1)] \tag{9}$$

The measurement vector is

$$\underline{y}(n) = \underline{G}\underline{x}(n) + \underline{v}(n) \tag{10}$$

Where $\underline{v}(n)$ is zero mean, white Gaussian noise with covariance $\Re_c$.

$$\Re_c = E\{\underline{y}(n).\underline{y}^t(n)\} \tag{11}$$

The gain (weights) vector is dynamically computed as

$$\underline{K}(n) = \underline{P}(n \mid n-1)\underline{G}^t[\underline{G}\underline{P}(n \mid n-1)\underline{G}^t + \Re_c]^{-1} \tag{12}$$

Where the measurement noise matrix $\underline{P}$ represents the predictor covariance matrix, and is equal to

$$\underline{P}(n+1 \mid n) = E\{\underline{x}_s(n+1)\underline{x}_s{}^*(n)\} = \underline{\Phi}\,\underline{P}(n \mid n)\underline{\Phi}^t + \underline{Q}] \tag{13}$$

Where $\underline{Q}$ is the covariance matrix for the input $\underline{u}$,

$$\underline{Q} = E\{\underline{u}(n).\underline{u}^t(n)\} \tag{14}$$

The corrector equation (covariance of the smoothed estimate) is

$$\underline{P}(n \mid n) = [\underline{I} - \underline{K}(n)\underline{G}]\underline{P}(n \mid n-1)] \tag{15}$$

Finally, the predictor equation is

$$\underline{x}(n+1 \mid n) = \underline{\Phi}\underline{x}(n \mid n) \tag{16}$$

A. The Singer-Kalman Filter

The Singer filter is a special case of the Kalman where the filter is governed by a specified target dynamic model whose acceleration is a random process with autocorrelation function given by

$$E\{\ddot{x}(t)\ddot{x}(t+t_1)\} = \sigma_a^2 e^{\frac{-|t_1|}{\tau_m}} \tag{17}$$

Where τ_m is the correlation time of the acceleration due to target maneuvering or atmospheric turbulence. The correlation time τ_m may vary from as low as 10 seconds for aggressive maneuvering to as large as 60 seconds for lazy maneuvering cases.

The maneuvering correlation coefficient is given by

$$\rho_m = e^{-\frac{T}{\tau_m}} \tag{18}$$

The transition matrix that corresponds to the Singer filter is given by

$$\Phi = \begin{bmatrix} 1 & T & \frac{1}{\beta_m^2}(-1+\beta_m T+\rho_m) \\ 0 & 1 & \frac{1}{\beta_m}(1-\rho_m) \\ 0 & 0 & \rho_m \end{bmatrix} \tag{19}$$

Where β_m is equal to $1/\tau_m$.

Typically, the sampling interval T is much less than the maneuvering time constant τ_m; hence, Eq. (19) can be accurately replaced by its second order approximation. More precisely,

$$\Phi = \begin{bmatrix} 1 & T & \frac{T^2}{2} \\ 0 & 1 & T(1-\frac{T}{2\tau_m}) \\ 0 & 0 & \rho_m \end{bmatrix} \tag{20}$$

For the short sampling interval case ($T << \tau_m$), the state transition matrix is computed from Eq. (20) as

$$\lim_{\beta_m T \to 0} \Phi = \frac{2\sigma_m^2}{\tau_m} \begin{bmatrix} 1 & T & \frac{T^2}{2} \\ 0 & 1 & T \\ 0 & 0 & 1 \end{bmatrix} \tag{21}$$

The covariance matrix is given by

$$\lim_{\beta_m T \to 0} \underline{C} = \frac{2\sigma_m^2}{\tau_m} \begin{bmatrix} \frac{T^5}{20} & \frac{T^4}{8} & \frac{T^3}{6} \\ \frac{T^4}{8} & \frac{T^3}{3} & \frac{T^2}{2} \\ \frac{T^3}{6} & \frac{T^2}{2} & T \end{bmatrix} \tag{22}$$

Where σ_m is the maneuver standard deviation.

While the Kalman filter is known to produce an optimal estimate of the target state when given the sequence of sensor measurements corrupted with white Gaussian errors, the computational burden of maintaining the Kalman filter may limit its use when many targets are being tracked. [2]

V. Comparison Between Performance of $\alpha\beta\gamma$ and Kalman Filter

To compare the performance of Kalman and αβγ filter we have considered them to track the same target with specified noise variance ($\sigma_v^2 = 0.5$). Both the cases of slow and fast targets have been observed.

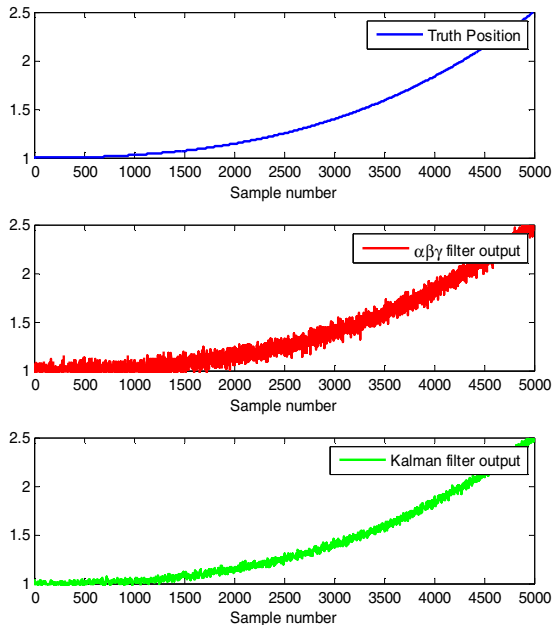


Fig. 4. True and Predicted positions from both filters (slow target).

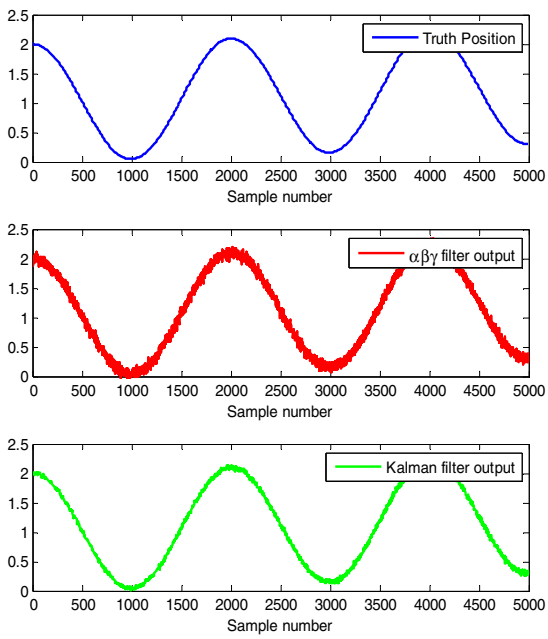


Fig. 5. True and Predicted positions from both filters (fast target).

Next, we vary noise variance and plot the mean residual error with respect to varying noise variance. Residual error from both the tracking filters are plotted on the same plot for better comparison. Both the cases of lazy maneuvering (slow target) and aggressive maneuvering (fast target) targets have been shown.

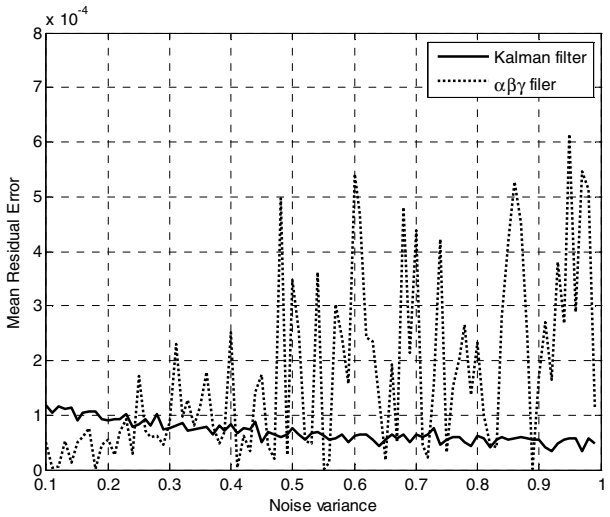


Fig. 6. Noise variance vs mean residual error (Slow target with noise).

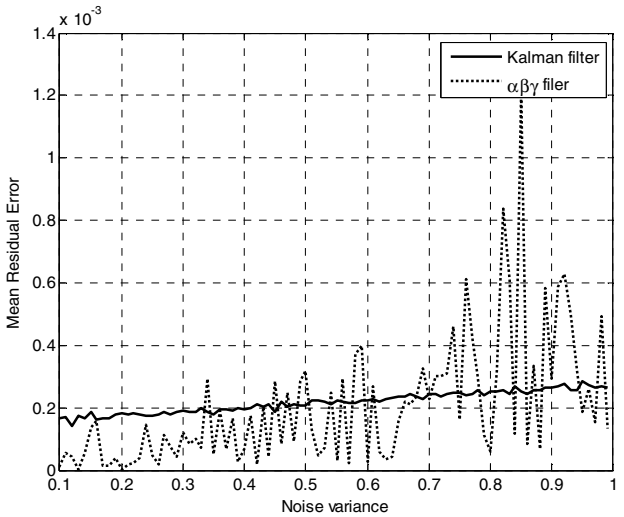


Fig. 7. Noise variance vs mean residual error (Fast target with noise).

In Fig. 4 ~ 5 a specified noise variance has been considered for both targets ($\sigma_v^2 = 0.5$). It is clearly seen that Kalman filter could track the targets with higher accuracy than αβγ filter. In Fig. 6 ~ 7 mean residual error from both αβγ filter and Kalman filter has been plotted on the same axis with varying noise variance. The plots show that with increasing noise variance residual error increases in case of αβγ filter but remains fairly constant in case of Kalman filter.

VI. CONCLUSIONS

By instructing both the constant coefficient αβγ filter and variable coefficient Kalman filter to track the same target we observe that Kalman filter gives more accurate tracking prediction than αβγ filter. With higher noise variance the performance of αβγ filter deteriorates but Kalman filter works on the basis of varying its gain coefficients in order to minimize the mean-squared error. Thus Kalman filter could track the targets with moderate accuracy even with high noise variance. Mean residual error vs noise variance plots reveal that with increasing noise variance residual error increases in case of αβγ filter but remains fairly constant in case of Kalman filter. This evidently shows the superiority of variable coefficient Kalman filter over constant coefficient αβγ filter.

Variable coefficient tracking filters can track objects with more accuracy than constant coefficient filters in expense of higher computational load. Consequently, before choosing an estimation filter, most probable target dynamics, hardware resources and acceptable error level should be investigated. An estimation filter which matches these requirements will be 'the best estimation filter'.

REFERENCES

[1] Mahafza, B. R., Radar System Analysis and Design; Chapman & Hall/CRC, 2004.

[2] Koray Akçay, "Performance metrics for fundamental estimation filters", Thesis submitted to the Graduate School of Natural And Applied Sciences, Middle East Technical University, Ankara, Turkey.

[3] Skolnik, M. I., Introduction to Radar Systems, 3rd edition; McGraw-Hill, Publishing Company, New York, 2008.

[4] Tait P., Introduction to Radar Target Recognition; The Institution of Engineering and Technology, 2005.

[5] Electronic Warfare and Radar Systems Engineering Handbook; Avionics Department AIR-4.5 Washington, DC 20361

[6] Eli Brookner, Tracking and Kalman Filtering Made Easy, John Wiley and Sons Inc., 1998.

[7] Tenne D., Singh, T., "Characterizing performance of α-β-γ filters", IEEE Transactions on Aerospace and Electronic Systems, Vol - 38, Issue - 3; July 2002.

Mobile Phone Based Telemedicine Service for Rural Bangladesh: ECG

Umme Sayma Busra
Dept of Computer Science and Engineering
Jahangirnagar University
Savar ,Dhaka, Bangladesh
saymabushra@gmail.com

Mohammad Zahidur Rahman
Dept of Computer Science and Engineering
Jahangirnagar University
Savar ,Dhaka,Bangladesh
rmzahid62@gmail.com

***Abstract*— Cardiovascular diseases such as; heart attack (coronary thrombosis, myocardial infarction), hypertension, stroke, coronary artery disease (CAD), congestive heart failure, peripheral artery disease (PAD) and diabetes create a new concern worldwide due to its harmness. In Bangladesh where majority live in rural areas that lack specialist care, we envision the need for much larger Internet-based telemedicine systems that would enable a large pool of doctors and hospitals to collectively provide healthcare services to entire populations. We propose a scalable, Internet-based architecture for telemedicine (ECG) to integrating multiple hospitals and mobile medical specialists focusing the rural health centers. This system based on the push model. For the necessity of the research at first we conduct a short survey among different health professional to justify the study more effectively.**

***Keywords*--ECG, Push model, EKG shield, Database server, ECG measuring units**

I. INTRODUCTION

Lifestyle diseases such as heart attack (coronary thrombosis**,** myocardial infarction), hypertension**,** stroke, coronary artery disease (CAD**),** congestive heart failure, peripheral artery disease (PAD) and diabetes lead to arteriosclerosis, which is one of the risk factors for developing cardiac diseases and prevention and early detection have become one of the most critical issues [1-4]. To prevent or limit the consequences of lifestyle diseases and promote health condition among middle-aged and elderly people, a health management system providing effective exercise prescriptions is required. Since effective and safe prescriptions should be based on evidence from physical data, it is necessary to acquire such daily physical data and vital signs and to evaluate the suitability of this physical information for each user. Therefore, a mobile medical system to manage and evaluate daily physical activities and to provide effective exercise prescriptions is desirable [5].

However, in developing nations like Bangladesh where majority live in rural areas that lack specialist care, we envision the need for much larger Internet-based telemedicine systems that would enable a large pool of doctors and hospitals to collectively provide healthcare services tothe entire population. The primary health care providers in the rural areas are mostly inexperienced [6]. In Bangladesh, about 70% of the total population lives in the rural areas. But around 75% of the country's doctor have intention to practice in the urban settings [7]. Overall, there is only one doctor for every 850 people [8]. However, the telecommunications infrastructure in Bangladesh is continuously improving. All the above-mentioned facts make telemedicine an appropriate technology for developing countries, particularly Bangladesh.

Hence our main objective is to develop a scalable, Internet-based architecture for telemedicine (ECG) to integrate multiple hospitals and mobile medical specialists focusing the rural health centers.

Telemedicine is defined as the use of telecommunication to provide medical information and services remotely. By the definition, telemedicine is the practice of health communication using audio, visual and data [9].Multimedia Messaging Service (MMS) is one of the communication applications that provide message sending services in many formats, like text, photo, graphic, animation, slide presentation, voice or video clip. Many new technologies are combindly used with browser, server, and new markup language in the modern mobile devices [10]. Portable Vital Sensing home medical system to manage and evaluate daily physical activities and to provide effective exercise prescriptions is desirable[11].Grid-Based Approach to Mobile Telemedicine is a scalable, Internet-based P2P architecture for telemedicine integrating multiple hospitals, mobile medical specialists, and rural mobile units. [12], [13].

The remainder of the paper is organized as follows: section 2 provides the rational of the study. The research Methodology is described in section 3. Description of system model is discussed in section 4. Section 5 provides the system requirement. ECG monitoring system design are discussed in section 6.System configuration are described in section 7 and system outcome are given in section 8. Section 9 describes the cost of health care. Finally challenges are described in section 10 and the conclusion is drawn in section 11.

II. RATIONAL OF THE STUDY

By the research findings an effective treatment of cardiovascular diseases in rural areas will develop. It's known to all;that in Bangladesh a large portion of the population lives in rural settings. And as the inhabitants of a developing country,where most of the people are still living under the livelihood of less than 1$, it's quite impossible for them to come in contact of good treatment. Sometime due to lack of

medical equipment in the remote areas,they are totally helpless to identify their diseases. This is why we should consider choosing the blessing of modern tele-medicine system for the rural setting. Because in Bangladesh till now,the traditional analog methods are used, this is costly enough. Our proposed services will not only user friendly but also affordable.

From the short survey we justify the rational of the research [14]. Based on different health professionals' feedback we justified our research rational;in addition to perceive in need for the services from previous works. Among them most of them were preferred for the ECG services [14].

TABLE I. TYPE OF MEDICAL TECHNOLOGY MOSTLY PREFRRRED BY DIFFERENT HEALTH PROFESSIONAL

Medical services	Responses		Percent of Cases
	N	percent	
X-ray	19	17.0%	82.6%
Blood pressure meter	21	18.8%	91.3%
Gluco Meter	20	17.9%	87.0%
Pulse Oximeter	9	8.0%	39.1%
Ultrasonograph	21	18.8%	91.3%
ECG	22	19.6%	95.7%
Total	112	100.0%	487.0%

III. METHODOLOGY

It was an Experimental research in design. At first we conducted a survey within 30 physicians by using simple random sampling method. Results from the survey lead us for the unmet need for ECG services by using Tele-medicine technology. The surveys were conducted in Dhaka. Vast numbers of our surveyed physicians were first experienced to work at rural settings. We have conducted experiment of our model in laboratory environment only.

IV. SYSTEM MODEL

In this research, we attempted to design a telemedicine system, especially for tele-diagnosis, using medical toolkit and android/Java technology for rural medical centers of Bangladesh. For this proposed system;here we have to use a medical toolkit, which is capable of measuring daily health conditions of electrocardiogram (ECG), and this digital signal is transfered to a receiving device (ECG Circuit) for signal processing and we use EKG shield for converting them as usable for 3G/2G network. Then this digital signal or data will be used for laptop/ Android Phone and it will be sent to another laptop/Android Phone using Internet .And when the receiver will receive this signal with showing the curve, the doctor can diagnosis the disease.

A. Proposed ECG System model

This Internet-based scalable system integrates multiple hospitals, mobile medical specialists and rural mobile units/clinics to form a large virtual enterprise. The current version of the telemedicine system is based on the push model—the biomedical signals of the patient are stored in the data base and then the signal is sent to an appropriate doctor.
Health workers carry ECG circuit device for measuring ECG. These ECG circuit devices are connected with a Smartphone/Laptop/Notebook, which uploads the data into the rural medical center database server. The medical center sends the request through internet connection to the nearest available doctor. The doctor then makes an evaluation of the measured data and provides consultations by telephone/Internet.

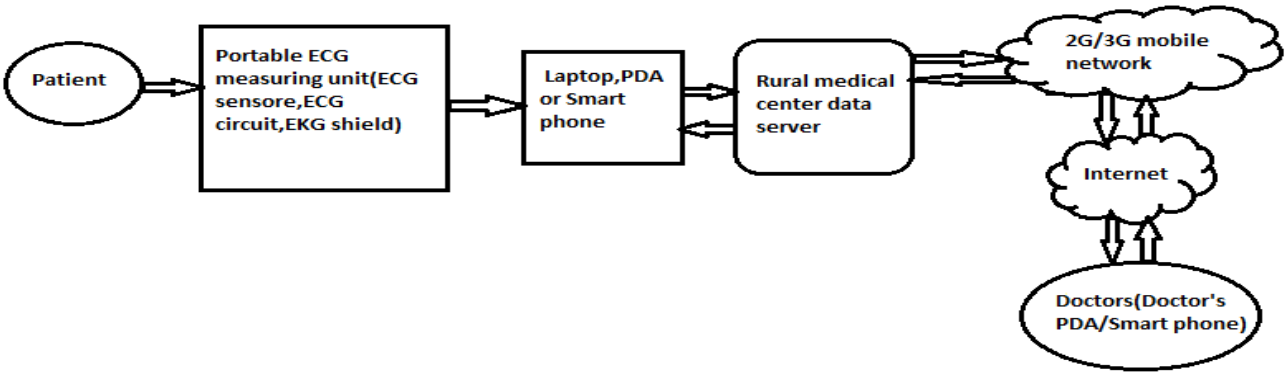


Fig.1: Telemedicine System for Electrocardiogram (ECG)

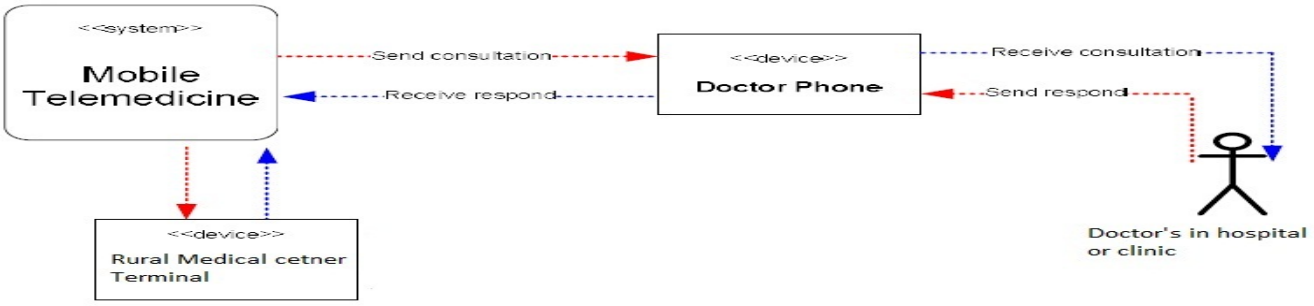


Fig. 2: Context diagram of the telemedicine system for ECG of rural medical center.

V. SYSTEM REQUIREMENT

For implementation,the system prerequisites are: 1) Portable ECG measuring Units, 2) Rural Medical Server, 3) Rural Medical Terminal, 4) Rural Medical Destination Terminal, 5) ECG Protocol Medical Data for Health Monitoring

1) Portable ECG measuring Units: The portable ECG measuring units consist of ECG sensor, ECG circuit and arduino EKG shield.ECG sensor measures the pulse wave and EKG shield converts the pulse wave to binary data for usability of 3G network.

2) Rural Medical Server: The rural medical server consists of a small computer. And the rural medical server has a wireless-LAN module to connect with Internet and a LAN module to communicate portable ECG measuring units. The server can transfer the measured data to a remote doctor's mobile,When the doctor's mobile is connected to the server .In server side we get the binary data from portable ECG measuring units. After that;a ECG curve is created by using RRD chart and saved in the data base table.

3) Rural Medical Terminal: A rural medical terminal (personal laptop, Internet enabled PDA or smart phone) is installed in the rural medical center. The rural medical center is connected with the rural medical server by LAN.

4) Rural Medical Destination Terminal: A rural medical destination terminal is the doctor's personal laptop, Internet enabled PDA or smart phone .This rural medical destination terminal is connected with the rural medical server by 3G internet network.

5) ECG Protocol Medical Data for Health Monitoring: ECG protocol medical data for health monitoring is used;in order to handle more information about patients and their vital signs.Which are asked by doctors through a open ended question.Which are the primary necessary information for patient health monitoring[14].

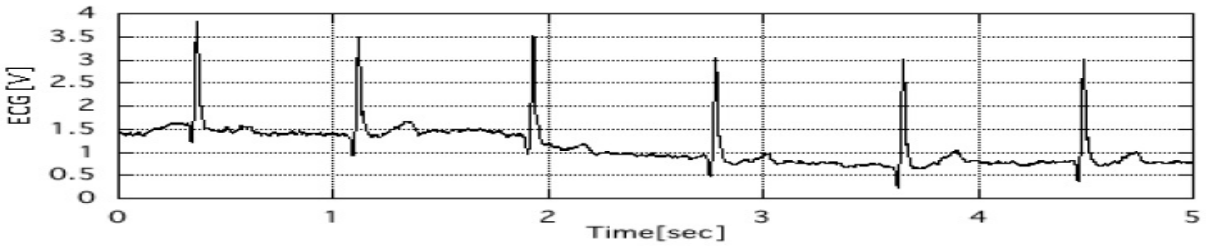


Fig. 3: The waveform of ECG measured by Binary data.

TABLE II. PATIENT DATA/ECG ACQUISITION DATA

ID	Contents	Length	Value(Parameter Data)
1	Name	1-15	String
2	Age	1-3	Number
3	Weight	1-3	Number
4	Height	1-3	Number
5	Patient ID	1-5	Number
6	Drugs	200	String
7	Heart Rate	20	Alphanumeric
8	Blood Pressure	30	Alphanumeric
9	ETT(exercise tolerance text)	300	Alphanumeric
10	Echo(2D & 3D mood) cardiac grapy	300	Alphanumeric
11	Cardiac Angiogram	300	Alphanumeric
12	Operational Mood	1	(Binary)[Basic(0), Emergency(1)]
13	Cardiac Enzyme –troponiu I & CK.MB (creatine kinase, myocardial band)	300	Alphanumeric
14	Medical Allergy or Drug Allergy	3-1000	(Binary) *Byte* *Contents* 1 binary: set equal to 255 0-Unspecified, 1-Penicillin,100 others
15	Food Allergy	3-1000	(Binary)
16	SPO2(saturation of partial oxygen)	50	Alphanumeric

TABLE III. DATA OF GENERATE ECG SIGNAL

ID	Contents	Value(Parameter data)
1	Patient ID	Number
2	Voltage(Heart muscles generate different voltages)	Binary
3	Time(sec)	Time

VI. ECG MONITORING SYSTEM DESIGN

The monitoring system of ECG for rural medical center is a push system .When the rural Medical Center Server finds new data (ECG) of a patient,then it sets its status as "P"; which means that the data is pending in database. The data is pushed to the client's side. Client side consists of doctor's mobile phone android apps. This is called TeleMed apps. And when the doctor opens the TeleMed apps, he/she can find the patient in queue. Doctors can check patient data and ECG curve through this TeleMed apps and diagnose the disease. After the doctor checks the patient ID, then this ID is removed from the patient queue and also notified to the server side, and sets the patient's status as "D"; that means done in the database.

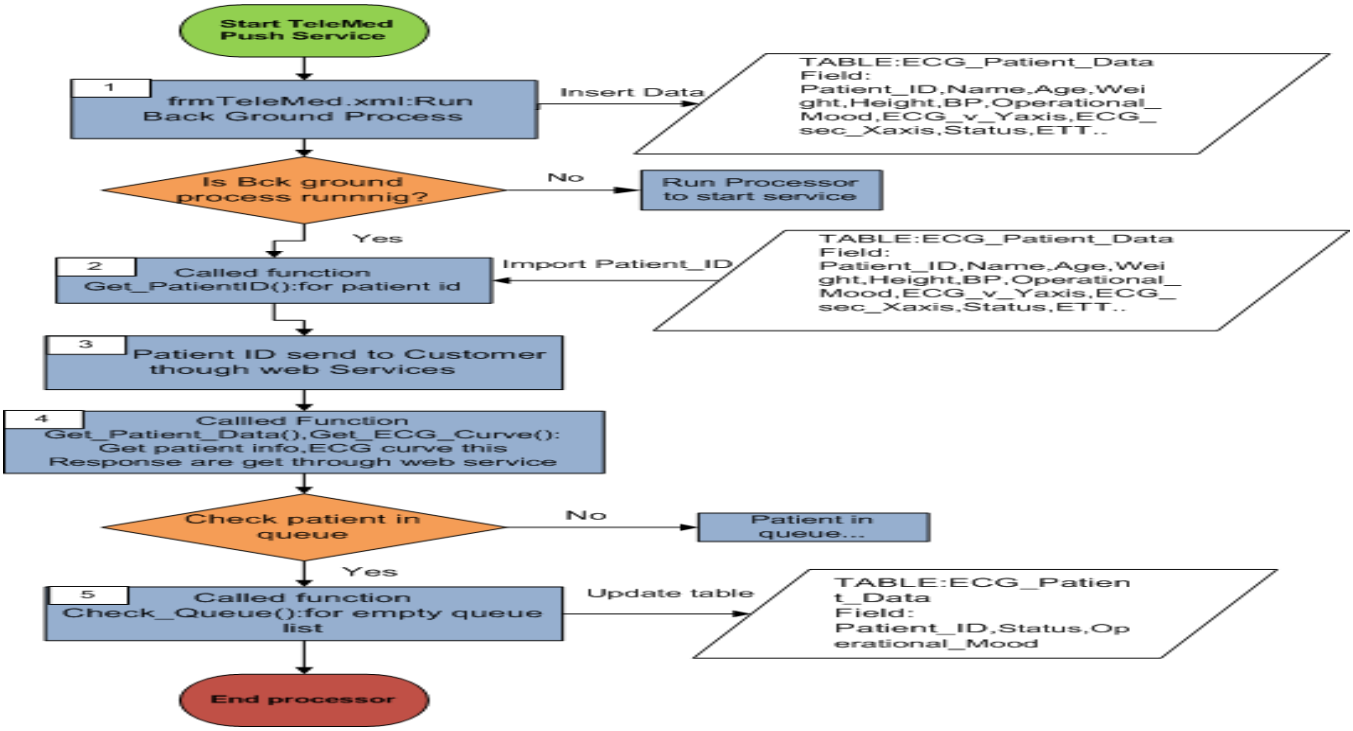


Fig. 4: System DFD.

VII. CONFIGURATION OF RURAL MEDICAL CENTER

The proposed rural medical system consists of three main components including a rural medical server, portable ECG measuring units (ECG sensor, ECG Circuit, Arduino EKG Shield) and a rural medical terminal. A rural medical server can accquire both physiological data and physical data measured by respective portable ECG sensor with ECG circuit and EKG shield, and can evaluate the patients' physiological conditions and convert the signal as binary data format. Each ECG measuring unit can be connected to the rural medical terminal. A rural medical terminal (personal laptop, Internet enabled PDA or smart phone) is installed in rural medical center .This rural medical terminal is connected with the rural medical server by LAN.

VIII. SYSTEM OUTCOME

If new patients are in database, doctor can see the total number of patient's details through the "telemed apps".Fig.5 shows the patient details where at first stage Fig.5(a) shows the number of patients .Patient ECG acquisition data could be seen in next stage at Fig.5(b) and the last stage at Fig.5(c) shows the patient ECG curve. When doctor checks the patient it will be automatically removed from the queue.

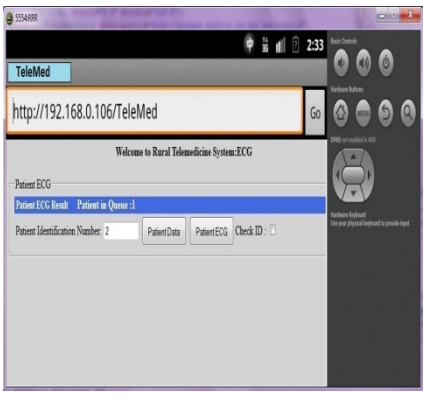


Fig. 5(a): System screen shot of patient in queue .

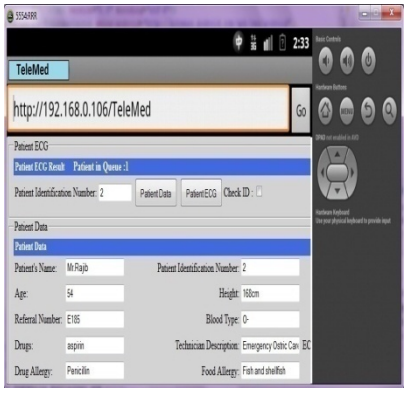


Fig. 5(b): System screen shot of patient data .

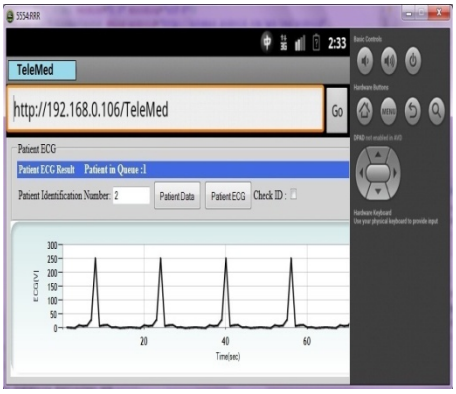


Fig. 5(c): System screen shot of patient ECG .

IX. TELEMEDICINE SYSTEM (ECG) AND THE COST OF HEALTH CARE

Telemedicine enables effective medical care delivery at a reduced cost compared to traditional, face-to-face care. Telemedicine offers the potential to lower fixed costs, by using less practitioners more efficiently. However, short term costs may in part of offset by:

1) Making the correct diagnosis early and avoiding costly complications (the notion of preventive medicine);

2) Making the correct diagnosis initially through a more focused approach;

3) Decreasing hospital length-of-stay through both quicker diagnosis and treatment.

4) Decreased transportation costs.

X. CHALLENGES

Challenges of Mobile Phone Based Telemedicine service for Rural Bangladesh: ECG to the consumer, provider and rural medical center. Table 4 shows the system challenges.

TABLE IV. CHALLENGES OF SYSTEM

Consumers	Providers	Rural Medical Center
Lowers level of trust between patient and provider	Requires a new design on the risk management	Poor internet facilities, network problem, no easily accessible.
Users may be intimidated by the technology (such as the elderly population)	Clinicians with superior expertise in ethics and practice may lack technical expertise or perspective	Insufficient electricity or generator services, limited technological support in rural areas.
The uncertainty of their medical consultation privacy	Lack of a relationship between patient and provider	Unwilling to expert professional to stay in rural area, unavailable of IT expert in rural areas.
More faith on traditional technology	May have difficulty in communicating with seniors who have an increase in dementia and sensory impairment	Requires to develop a telemedicine specific policies and procedures, inadequate of awareness on IT.

XI. CONCLUSION

A mobile telemedicine system for rural medical center is a network-based and distributed information system to connect rural patients to medical specialists. Therefore, it indicates to consider inter-operability of telemedicine systems in rural area. Regarding device connectivity to the rural medical server, the portable ECG measuring units have the advantage of being easily connected to a rural terminal and that composed in the host medical server. However considering network connectivity between a rural medical server and a remote data server, more sophisticated system architecture is required in terms of data formats and communication protocols. Our future concentration will be on trying to develop home medical server to establish home mobile telemedicine system.

ACKNOWLEDGMENT

Authors were supported by the Department of Computer Science and Engineering, Jahangirnagar University, Savar, Dhaka, Bangladesh.

REFERENCES

[1] M. Matsuda, “Effect of Exercise and Physical Activity on Prevention of Arteriosclerosis – Special Reference to Arterial Distensibility–,” International Journal of Sport and Health Science, Vol. 4, pp.316-324, 2006.

[2] E. G. Lakatta, D. Levy, “Arterial and Cardiac Aging: Major Shareholders in Cardiovascular Disease Enterprises, Part I: Aging

Arteries: A "Set Up"for Vascular Disease" Circulation, Vol. 107, pp.139-146, Jan, 2003.
[3] S. S. Najjar, A. Scuteri, and E. G. Lakatta, "Arterial Aging Is It an Immutable Cardiovascular Risk Factor?," Hypertention, Vol. 46, p. 454-462, 2005.
[4] P. D. Tompson, D. Buchner, I. L Pina, G. J Balady, M. A. Williams, et al. "Exercise and physical Activity in the Prevention and Treatment of Atherosclerotic Cardiovascular Disease: A Statement From the Council on Clinical Cardiology(Subcommittee on Exercise, Rehabilitation, and Prevention) and the Council on Nutrition, Physical Activity, and Metabolism (Subcommittee on Physical Activity),"Circulation, Vol. 107, pp.3109-3116, 2003
[5] "The Council on Clinical Cardiology (Subcommittee on Exercise, Rehabilitation, and Prevention) and the Council on Nutrition, Physical Activity, and Metabolism (Subcommittee on Physical Activity)," Circulation, Vol. 107, pp.3109-3116, 2003.
[6] "Research and technology development on telematics systems in health care," Annual technical report on RTD in health care, CEC DG XIII, AIM, Brussels, Belgium, 1993.
[7] http://www.tradingeconomics.com/bangladesh/rural-population-percent-of-total-popul ation -wb-data.html (Accessed on: October 22, 2012).
[8] http://www.populstat.info/Asia/bangladg.htm (Accessed on: October 25, 2012).
[9] Andik Setyono1, Md. Jahangir Alam2, and Raed Ali Al-Saqour3," Design of Multimedia Messaging Service for Mobile Telemedicine System ", International Journal of Network and Mobile Technologies ISSN 1832-6758 Electronic Version VOL 1 / ISSUE 1 / JUNE 2010.
[10] Edy Djunaedi , "Design and Implementation of Telemedicine Information System based on MMS Technology", stttelkom.ac.id,2006.
[11] F. Ichihashi1 and Y. Sankai," Development of a Portable Vital Sensing System for Home Telemedicine" Based on the proceedings of the 29th Annual International Conference of the IEEE EMBS Cité Internationale, Lyon, France August 23-26, 2007 .
[12] Center for Connected-Health at Partners Healthcare, Massachusetts General Hospital, Boston, Mass, USA, http://www.connected-health.org/ (Accessed on: November 25, 2012).
[13] BELGIUM-HF, "the largest Belgian telemedicine clinical trial on congestive heart failure," http://www.belgium-hf.be/. (Accessed on: November 25, 2012).
[14] Umme Sayma Busra,, "Mobile Phone based Telemedicine Service for Rural Bangladesh:ECG", MSc. thesis, Dept of CSE, Jahangirnagor University, 2013.

Budget Allocation Model for the Academic Library Acquisition Using Data Mining Technique

Md. Hossain, Rashedur M Rahman

Department of Electrical Engineering and Computer Science,
North South University
Plot-15, Block-B, Bashundhara, Dhaka 1229, Bangladesh
myhossain@yahoo.com, rashedur@northsouth.edu

Abstract— **Allocation of budget for acquisition of an academic library is an important initiative to enhance the efficiency of management, to accomplish the objectives ensuring the availability of materials in a library. Processing the circulation data of library, two important dimensions, namely, concentration and connection [1] could be explored among departments and library members. This can make the analysis easier by calculating weights in those two important dimensions to make the decision about budget allocation. This paper analyses the circulation data of North South University and suggests for efficient management and budget allocation of the North South University Library by using the above mentioned metrics.**

***Keywords*—Budget Acquisitions; Circulation Data; Utilization; Knowledge discovery.**

I. INTRODUCTION

Data Mining is a fabulous technique to acquire knowledge from any archive of scatterd data. It can easily help us taking any decision based on the collected knowledge. Allocation of budget for a library is important to manage and maintain it properly. Besides, good allocation ensures the maximum and optimum usability of the library materials for the sake of the members of the library. In this occasion data mining can play a vital role to allocate a perfect or nearly perfect budget for a library based on the circulation data. There are a number of research studies reported in the literature that used data mining techniques in the last few years. This paper introduces some easier techniques of statistical analysis and goal programming to extract the knowledge which can give us the power to take the decision about the allocation of budget. In this reasearch we analyse NSU Library Database for possible allocation of academic library acquisition budget via circulation database mining. We derived weights of acquisition budget allocation by composing the descriptive knowledge via utilization concentration and the suitability via utilization connection[1] for departments concerned. The value of concentration and connection is calculated from the number of records, the distribution of material categories used, and the relation between categories and subjects. We obtained the degree of concentration for a department by measuring information entropy using ID3 algorithm introduced by Quinlan [4]. Finally we obtain budget allocation table for each department using corresponding weights.

The rest of the paper is organized as below. Section II highlights some related papers that describe similar kind of researches. Section III shows the diagram of the database model, Section IV describes the model in detail. Section V presents our research findings and finally Section VI draws the conclusion.

II. RELATED WORKS

Many research works have been involved into budget allocation for library materials acquisition. Most of them are in common style and with a common objective, some have slightly different in objectives. Some of them are discussed here.

Kao, Chang and Lin [1] worked with the circulation database of Kun Shan University of Technology where they used ID3 [4] algorithm of Data Mining. Wu, Lee and Kao [2] pretended circulation statistics mechanism and an association rule was used to discover knowledge. There are several models approached by several researchers like ABAMDM [1], $KDBM_{LMA}$ [2] to help derive the utilization of library material categories. Although many approaches and research reports have been extensively used to help library material acquisitions, the knowledge contained in circulation databases has very rarely been used to investigate in-depth how the acquired materials are being used. Wu and Lee [3] presented a decision support model for library acquisition budget allocation using the knowledge derived from circulation databases.

A good research work was done by Rahman et al.[5] focused on three important goals : i) A balance between continuing commitment and new initiatives, ii) A balance between resources to support undergraduate learning and those to support graduate work and research, and iii) A balance between subject disciplines. In order to achieve these goals, the specific objectives of the study were to firstly, construct a mathematical model to determine the total amount of the library's budget that should be allocated to each faculty and secondly, construct a mathematical model for faculty's budget allocation in terms of acquisition of books and journals. The model must satisfy the budget allocated for every faculty and at the same time avoid overspending on the purchasing of materials.

Another analytical work by C. H Wu [6] reports the most beneficial use of the allocated acquisition budget. Similar to their approach we have also explored knowledge in the circulation databases in-depth to relevantly reflect the

need of budget allocation. In this research a data mining based model has been introduced based on the feature of ID3 algorithm to explore explanatory knowledge via information theory and statistics to derive appropriateness via utilization gain.

III. DATABASE MODEL

A. *Modified NSU Library Database*

We modified original NSU Library Database for our research as follows (presented in Table I):

TABLE I. NSU LIBRARY DATABASE

Original NSU Library Database	Modified Database for our research
1.books	materials (material_id,call_number,category_code)
2. client	dept_member (member_id, department_id)
3. department	department(department_id, department_name)
4. loanarchive	circulation (material_id, member_id)
5. there is no table for material_category	*Newly created table for material_category(category_code,description)

* There was no table for material_category. We created it with the help of Library of Congress Classification(*Reference:http://en.wikipedia.org/wiki/Library_of_Congress_Classification*). Portion of the table is shown in Table II.

TABLE II. MATERIAL_CATEGORY TABLE

id	category_code	description
1	A	General Works
2	AC	Collections. Series. Collected works
3	AE	Encyclopedias
4	AG	Dictionaries and other general reference works
5	AI	Indexes
6	AM	Museums. Collectors and collecting
7	AN	Newspapers
8	AP	Periodicals
9	AS	Academies and learned societies
10	AY	Yearbooks. Almanacs. Directories[disambiguation ne...

In original NSU books table (we named the books table in our database as materials), there was not category_code attribute. We generated the value for this attribute from call_number. The value of the category_code attribute is the first alphabetic digit parts before starting of numeric digits of a call_number which coincides with category_codes described in Library of Congress Classification. For example, if call_number of any material is HD9940.B36Q83 331.4 Q2e 2000, then category_code of this material is HD which means (from Library of Congress Classification) as "Industries, Land use, Labor" that we have stored in material_category table.

Mebership table is created by department and material_category tables which contains all departments, categories, and corresponding semantic strengths (we called it matching_level). The semantic strength can be expressed by the degree of the relation between a department and a material category. These values are management definable [1]. There are five levels for semantic strength, SS(A, H, M, L, N)—A: absolutely matching, H: highly matching, M: matching, L: likely matching, N: absolutely not matching. The corresponding importance is, IS(0.4, 0.3, 0.2, 0.1, 0.0).

Membership table with department_id='EECS' and not matching_level='N' is presented in Table III.

TABLE III. CATEGORY MATCHING LEVEL (MEMBERSHIP TABLE)

id	department_id	category_code	matching_level
13	EECS	BC	H
15	EECS	BF	L
61	EECS	GA	L
64	EECS	GE	L
65	EECS	GF	L
70	EECS	H	L
71	EECS	HA	M
72	EECS	HB	L
73	EECS	HC	L
74	EECS	HD	L

And we created level table which describes importance of each level which is shown in Table IV.

TABLE IV: LEVEL WITH IMPORTANCE (LEVEL TABLE)

level	importance
A	0.4
H	0.3
M	0.2
L	0.1
N	0.0

Category importance/strength graph for department EECS is shown in Fig. 1:

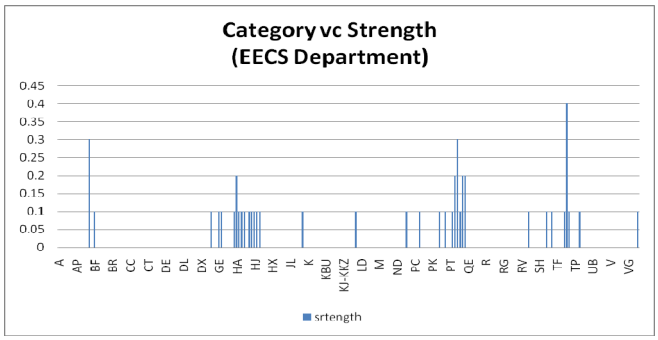


Fig. 1. Strength of different categories for EECS Dept.

B. *Structure of Database Model*

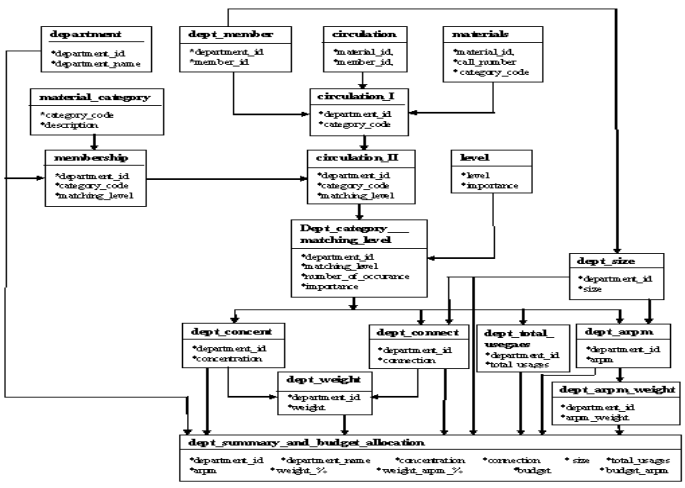


Fig. 2. Database model view

IV. MODEL DESCRIPTION

The architecture of our model is illustrated in Fig. 2. It contains two stages to achieve the objective of our research. The first stage is to pre-process the circulation data, and the second is to obtain descriptive knowledge and utilization gain, that are used to derive the weights of acquisition budget allocation for departments.

A. Pre process of circulation data

Circulation table is the heart of Library Database. From this table we generated circulation_I table linking with materials and dept_member tables. In above we already mentioned that we built a new table material_category table with the help of of Library of Congress Classification. Mebership table is created by department and material_category tables. So membership table has rows of (department size X category size). Actually it has attributes – department_id, category _code, matching_level (semantic strength). The semantic strength can be expressed as the degree of the relation between a department and a material category. These values are management definable. There are five levels for semantic strength, SS(A, H, M, L, N)—A: absolutely matching, H: highly matching, M: matching, L: likely matching, N: absolutely not matching. The corresponding importance is, IS(0.4, 0.3, 0.2, 0.1, 0.0). We created a level table which describes importance of each level.

Now from circulation_I and membership tables, circulation_II is generated. Here all records of original circulation table contained with attributes department_id, category_code and matching_level which is the semantic strength between a department and a category in membership table. After pre-process of circulation_II and linking the pre-processed table with level table, we got the important table dep_category_matching_level table which contains all department_id with total number of matching values and their relative importances. The table view of portion of this data is shown in Table V.

TABLE V. DEPARTMENTS WITH TOTAL NUMBER OF MATCHING LEVEL (DEP_CATEGORY_MATCHING_LEVEL TABLE)

department_id	matching_level	number_of_occurance	importance
ACT	A	7	0.4
ACT	L	2	0.1
ACT	N	8	0.0
ADMI	H	2	0.3

B. Obtaining descriptive knowledge and utilization gain

The descriptive knowledge is stored in table of dept_concent (department_id,concentration) and utilization gain in dept_connect (department_id,connection). Concentration represents how different categories of materials are used by the department. The less value of it indicates that the corresponding department use of materials in various subject. The more value of it means that it is pure and categories are well distributed towards the department. The connection represents the utilization suitability of a department. For example, the descriptive knowledge "EECS department used a good number of materials and most of them are in its subject" can relevantly explain the utilization of the EECS department.

Now we will obtain the concentration for a department. Applying the ID3 algorithm introduced by Quinlan (1986) with following formula[1], we can found the Information(I) of all departments and Entropy(E) of each department :

$$I\left(n_{C_1}, n_{C_2}, \ldots\ldots, n_{C_n}\right) = \left(-\frac{n_{C_1}}{M}\log_2\frac{n_{C_1}}{M}\right) + \ldots + \left(-\frac{n_{C_n}}{M}\log_2\frac{n_{C_m}}{M}\right) \quad (1)$$

n_{C_1} is the number of records that return to class C_i , $i=1,2,\ldots,n$ and M is the total number or records

$$E(D) = \sum_{i=1}^{t}\left[\left(\frac{n_{V_i}}{M}\right) I\left(a_{V_iC_1}, a_{V_iC_2}, \ldots a_{V_iCm}\right)\right] \quad (2)$$

t is the number of different values that the department D can takes on; n_{V_i} is the total number of records that the department D takes value V_i , $i=1, 2,\ldots,t$; $a_{V_iC_j}$ is the total number or records that the department D takes value V_i and returns to class C_j , $i=1, 2,\ldots,t$ and $j=1,2,3\ldots,m$ and M is the total number of records.

From above parameters and using following formula [1] , we can calculate concentration for each department :

$$\text{Concentration (D)} = I\left(n_{C_1}, n_{C_2}, \ldots\ldots, n_{C_n}\right) - E(D) \quad (3)$$

We wrote PHP programming (using dept_category _matching _ level table) to find out information I and Entropy by the above formula which is shown in Appendix.

In dept_concent table these values are stored. The table view of portion of this data is shown in Table VI.

TABLE VI. INFORMATION AND ENTROPY FOR DIFFERENT DEPARTMENT (DEPT_CONCENT TABLE)

department_id	information	entropy	concentration
ACT	2.0898	0.0029	0.0000
ADMI	2.0898	0.0124	0.0000
ARC	2.0898	0.0033	0.0000
BBA	2.0898	0.9673	0.0000

Then applying sql command, final concentration is stored in the Table VII.

TABLE VII. CONCENTRATION FOR DIFFERENT DEPARTMENTS (UPDATED DEPT_CONCENT TABLE)

department_id	information	entropy	concentration
ACT	2.0898	0.0029	2.0869
ADMI	2.0898	0.0124	2.0774
ARC	2.0898	0.0033	2.0865
BBA	2.0898	0.9673	1.1225

Now, to obtain the connection for a department, we created the dept_connect (department_id,connection) table by using dept_category _matching _ level table (presented in Table V) and following formula[1] :

$$\text{Connection(D)} = \frac{\sum_{i}^{n} n_{L_i}\omega_{L_i}}{N_D} \quad (4)$$

where, N_{L_i} is the number of records of which the matching_level is L_i and ω_{L_i} is the importance of L_i . N_D is

the number of members in a department that could be generated from dept_member table.

The table view of portion of this data is given in Table VIII as follows:

TABLE VIII. CONNECTION FOR DIFFERENT DEPARTMENTS (DEPARTMENT_ID,CONNECTION TABLE)

department_id	connection
ACT	0.1875
ADMI	0.0795
ARC	0.0490
BBA	0.1504

Weight of each department is calculated by following formula [1] :

$$\text{Weight (D)}=\frac{\alpha Concentraiton(D)+(1-\alpha)Connection(D)}{\sum_{i=1}^{m}[\alpha Concentraiton(D_i)+(1-\alpha)Connection(D_i)]} \quad (5)$$

It is an assumption that the importance of concentration α and connection $1-\alpha$ are management definable.

We assume this value as 0.3 in our calculation. In dept_weight table, we stored this value and shown in Table IX.

TABLE IX. WEIGHT FOR DIFFERENT DEPARTMENTS (DEPT_WEIGHT TABLE)

department_id	weight
ACT	0.0560
ADMI	0.0502
ARC	0.0488
BBA	0.0327

We also created dept_total_usages (department_id,total_usages) and dept_arpm(department_id, arpm) from dept_ category_matching_level table and then dept_arpm_weight(department_id, arpm_weight).

Finally we created dept_summary_and_budget_ allocation table, attributes of which are department_id, department_name, concentration, connection, size, total_usages, arpm, weight_%, weight_arpm_%, budget and budget_arpm.

The table view of left portion of this data is given in Table X and right portion in Table XI.

TABLE X. CONCENTRATION AND CONNECTION FOR DEPARTMENTS (LEFT PORTION OF DEPT_SUMMARY_AND_BUDGET_ALLOCATION TABLE)

department_id	department_name	concentration	connection	size	total_usages
ACT	Accounts	2.0869	0.1875	16	17
ADMI	Adminisration	2.0774	0.0795	78	63
ARC	Architecture	2.0865	0.0490	104	17
BBA	Bachelor of Business Administration	1.1225	0.1504	7950	4384
BIOT	Biotechnology	2.0851	0.0607	89	20
ECO	Economics	1.9832	0.1670	721	434
EECS	Electrical Engineering and Computer Science	1.7048	0.1930	2201	1605
ENG	English	2.0323	0.1619	373	283
ENV	Environmental Studies	2.0416	0.1769	234	179
ETE	Electronic and Telecommunication Engineering	1.9373	0.2433	515	564

TABLE XI: DIFFERENT BUDGET ALLOCATION SCHEME (RIGHT PORTION OF DEPT_SUMMARY_AND_BUDGET_ALLOCATION TABLE)

total_usages	arpm	weight_%	weight_arpm_%	budget	budget_arpm
17	1.0625	5.6000	7.4300	56000.0000	74300.0000
63	0.8077	5.0200	5.6400	50200.0000	56400.0000
17	0.1635	4.8800	1.1400	48800.0000	11400.0000
4384	0.5514	3.2700	3.8500	32699.9980	38500.0000
20	0.2247	4.9400	1.5700	49400.0000	15699.9990
434	0.6019	5.2600	4.2100	52600.0000	42100.0000
1605	0.7292	4.7800	5.1000	47800.0000	51000.0000
283	0.7587	5.3500	5.3000	53500.0000	53000.0000
179	0.7650	5.4500	5.3500	54500.0000	53500.0000
564	1.0951	5.5600	7.6500	55600.0000	76500.0000

V. RESULTS AND FINDINGS

By utilizing ID3 algorithm by formula (4) and formula (5), the results including concentration, connection, weights and Average Record Per Member (ARPM) and allocated acquisition budget for our analysis was shown in Table XII. To determine the budget amount for each depaertment total budget has been considered 1,00,00,000 BDT. Our budget % and ARPM % are percentages of this total budget amount. The results found by ARPM was also included in Table XII. The relationship between concentration and connection was shown in Fig. 3 while the number of members, the number of records, and allocated budget is shown in Fig. 4.

TABLE XII. THE RESULTS VIA OUR ANALYSIS AND ARPM (DEPT_SUMMARY_AND_BUDGET_ALLOCATION TABLE)

dept_id	department_name	concentration	connection	size	total_usage	arpm	weight_%	weight_arpm_%	budget	budget_arpm
ACT	Accounts	2.0869	0.1875	16	17	1.0625	5.6	7.43	56000	74300
ADMI	Adminisration	2.0774	0.0795	78	63	0.8077	5.02	5.64	50200	56400
ARC	Architecture	2.0865	0.049	104	17	0.1635	4.88	1.14	48800	11400
BBA	Bachelor of Business Admin	1.1225	0.1504	7950	4384	0.5514	3.27	3.85	32700	38500
BIOT	Biotechnology	2.0851	0.0607	89	20	0.2247	4.94	1.57	49400	15699.999
ECO	Economics	1.9832	0.167	721	434	0.6019	5.26	4.21	52600	42100
EECS	Electrical Engineering and C	1.7048	0.193	2201	1605	0.7292	4.78	5.1	47800	51000
ENG	English	2.0323	0.1619	373	283	0.7587	5.35	5.3	53500	53000
ENV	Environmental Studies	2.0416	0.1769	234	179	0.765	5.45	5.35	54500	53500
ETE	Electronic and Telecommur	1.9373	0.2433	515	564	1.0951	5.56	7.65	55600	76500
GCE	General & Continuing Educa	2.0875	0.0229	109	25	0.2294	4.75	1.6	47500	16000.001
IS	Information Services	2.0841	0.18	15	27	1.8	5.56	12.58	55600	125800
LIB	Library	2.0812	0.1033	30	114	3.8	5.15	26.56	51500	265600
LSC	Life Sciences	2.0891	0.0343	175	16	0.0914	4.81	0.64	48100	6400
MBA	Master of Business Adminis	2.0455	0.0892	759	226	0.2978	5	2.08	50000	20800
MDS	Master of Development Stu	2.0898	0.0923	26	6	0.2308	5.11	1.61	51100	16100.001
MPH	Masters in Public Health	2.0881	0.0701	147	28	0.1905	5	1.33	50000	13300
PHAR	Pharmacy	2.0683	0.1202	585	208	0.3556	5.21	2.49	52100	24900
REG	Registrar	2.0855	0.0079	114	55	0.4825	4.67	3.37	46700	33700
UKN	Unknown	2.0898	0	28	2	0.0714	4.64	0.5	46400	5000

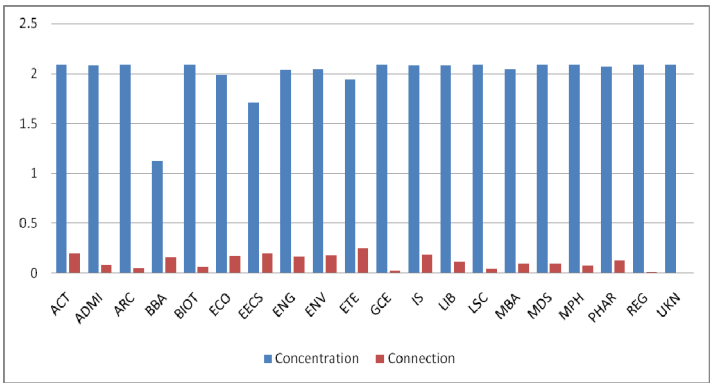


Fig. 3. Department wise Concentration-Connection Graph

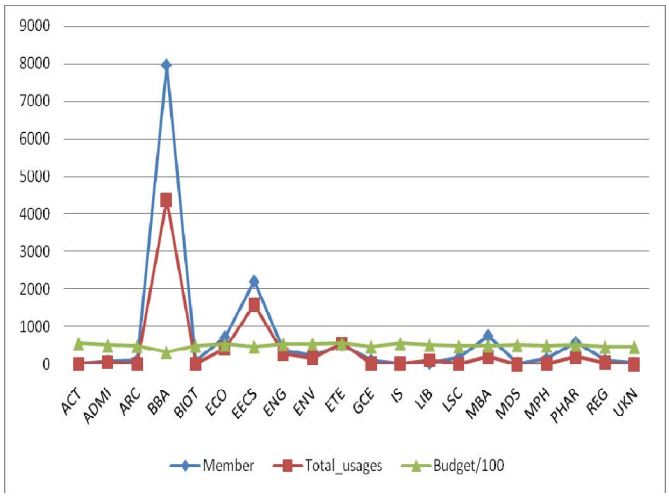


Fig. 4. Department vs Member, Total_usages and Final Budget (Budget/100)

To fit the graph in this small area, the values of the budget were devided by 100 in Fig. 4. From the tables and figures given above, it was found that:

1. Allthough the department of Electronic and Telecomunication Engineering (ETE) showed the highest ARPM (1.0951 in Table XII), it has obtained the highest connection but not the budget. Good connection and ARPM value here indicates that the memebers of EECS Department has made a good use of the materials and all of them are related to their areas.
2. In Fig. 3, the department of Master of Development Studies (MDS) showed the highest concentration (2.0898 in Table XII). This idicated that the categories used were equally distributed. However, the Fig. 3 indicated that it did not get a high value of connection. This indicated that most of the materials used were not in its subject. Fig. 4 supported this information and showed that it got the low budget.
3. In Fig. 3, the department of Bachelor of Business Administration (BBA) got the lowest value of concentration (1.1225 in Table XII), but the largest number of usage (4,384 in Table XII). Accordingly, it was shown that it used the materials in a variety of categories, which indicated that part of these materials were in its subject, but part were not. Therefore, the total connection it found was not high (in Fig. 3), and so was the allocated acquisition budget (in Fig. 4).
4. In Fig. 3, it was found that the department of ETE got the highest value of connection (0.2433 in Table XII), but the observed concentration was considered to be low (1.9373) in comparison to others (in Fig. 3) and the number of members was not high (515 in Table XII). However, it finally got almost highest allocated budget (55600.00 in Table XII). It can be said that the value of importance of concentration (α) took a very significant role in this case.
5. The acquisition budget allocated via ARPM depended totally upon the ARPM. Fig. 5 provided the information that the result produced by our analysis was fairly different from that by ARPM.

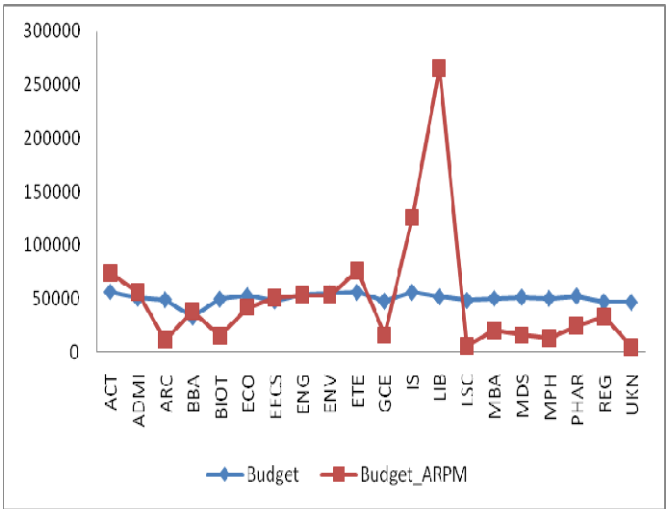


Fig. 5. The allocated acquisition budget via our analysis and ARPM for 20 departments

VI. Conclusions

In this paper we have mined the library circulation data to understand what budget should be allocated for the next academic year using ID3 algorithm. We employed SQL to help pre-process the circulation data if necessary. We also employed the information theory to measure the concentration of categories observed in the circulation data table, and at the same time the utilization connection to derive the weights as a decisional base of acquisition budget allocation with the help of goal programming using a high-level language. In our case we used PHP with MySQL database. The knowledge discovered by our analysis can be used to support making decisions in regard to the acquisition library budget allocation for North South University or any other university having the similar operation process. This knowledge will help in a great deal to allocate the budget for the library and will save a large amount of time and cost. It also helps to reduce mismanagement of library resources by ensuring appropriate availability of materials for every members of the library.

Finally, a core question remains while budgeting and that is how to capitalize the investment to buy right things. It will really be a ceaseless discovery of unknown information in historical data to support the decisions related to the allocation of budget.

ACKNOWLEDGMENT

First of all we would like to thank the Almighty Allah to make us capable of completing this research work with such ease. We also express our gratitude to the faculty members of EECS department of North South University to help us to collect bunch of information and references. Finally special thanks to the Library authority to help us by giving access and permission to continue this research work on their data.

REFERENCES

[1] C. Kao, H.C. Chang, C.H. Lin, "Decision support for the academic library acquisition budget allocation via circulation database mining." Information Processing & Management, Volume 39, Issue 1, pp. 133-147, December, 2003.

[2] C.H. Wu, T.Z. Lee, S.C. Kao, "Knowledge discovery applied to material acquisitions for libraries", Information Processing & Management, Volume 40, Issue 4, pp.709-725, July, 2004.

[3] C. H. Wu and T.Z. Lee, "Material Acquisitions Using Discovery Informatics Approach", Encyclopaedia of Data Warehousing and Mining, pp. 705-709, IGI Global, 2005, Web. 10 Nov. 2013. doi:10.4018/978-1-59140-557-3, ch134.

[4] J.R. Quinlan, "Induction of decision tree. Machine Learning", Volume 1, 81–106, 1986.

[5] S. A. Rahman and N.M. Yousop, "Modeling of Budget Allocation for University Library", Journal of Statistical Modeling and Analytics, Vol.2, No.2, pp. 1-8, 2011.

[6] C. H Wu et al., "Data mining applied to material acquisition budget allocation for libraries: design and development", Expert Systems with Applications, Volume 25, Issue 3, pp. 401–411, October, 2003.

APPENDIX: PDEUDO CODES OF PHP PROGRAMMING

A. id3_I.php

```
<?php
mysql_connect('localhost','root','');
mysql_select_db('library_new');
$result1 = mysql_query('select  sum(number_of_occurance)
as sum_occurance from  dept_category_matching_level ');
$row=mysql_fetch_array($result1);
$totalM=$row['sum_occurance'];
$result2 = mysql_query('select matching_level,
sum(number_of_occurance) as total_occurance from
dept_category_matching_level group by matching_level');
while($row = mysql_fetch_array($result2)){
	if($row['total_occurance'] <= 0)
		$I += 0;
	else
	$I+= -
($row['total_occurance']/$totalM)*log(($row['total_occuranc
e']/$totalM), 2);
	}
	mysql_query('update dept_concent set
information="'.$I.'" ');
?>
```

B. id3.php

```
<?php
mysql_connect('localhost','root','');
mysql_select_db('library_new');
$allM = array();
$nC = 0;
$C = array('A'=>0,'H'=>0,'M'=>0,'L'=>0,'N'=>0);
$I = $E = array();
$result = mysql_query('select department_id,
sum(number_of_occurance) as total_occurance from
dept_category_matching_level group by department_id');
while($row = mysql_fetch_array($result)){
	$allM[$row['department_id']] =
$row['total_occurance'];
}
// Find Information for every department
$totalM = 0;
foreach($allM as $dept=>$M){
	$deptCatMatching = mysql_query('select
matching_level, number_of_occurance from
dept_category_matching_level where department_id =
'".$dept."' ');
	while($rowMatching =
mysql_fetch_array($deptCatMatching)){
		if($rowMatching['number_of_occurance']
<= 0)
			$I[$dept] += 0;
		else
			$I[$dept] += -
($rowMatching['number_of_occurance']/$M)*log(($rowMatc
hing['number_of_occurance']/$M), 2);
	}
	$totalM += $M;
	}
foreach($allM as $dept=>$M){
	$E[$dept] += ($M/$totalM) * $I[$dept];
	echo "E of
".$dept.":   ".$E[$dept]."<br />";
	mysql_query('update dept_concent set
entropy="'.$E[$dept]'" where department_id="'.$dept.'"');
}
```

Low complexity Offline Character Recognition of License Plate Using Perspective View

Jainal Uddin
Dhaka, Bangladesh
jainal36@gmail.com

Md. Mahbubur Rahman
BUET
Dhaka, Bangladesh

Mohammad Rakib, Md. Jannatul Ferdous, Shariar Kabir, Md. Ali Azam

***Abstract*— License plate detection and character recognition have unveiled new possibilities and challenges in the field of intelligent transport system. Numerous algorithms have been proposed regarding license plate localization and tracking, character segmentation and character recognition. License plate character recognition is still an active area of research specially in terms of processing complexity. A handful of feature extraction algorithms including vertical and horizontal histogram, center of gravity, structural characteristics, modified edge maps, image projection, multi zoning, adaptive multi zoning, concavities measurement, gradient directional features, median gradient features, longest-run features and momentum features have been proposed. Neural network followed by ensemble classifier shows recognition rate in the high nineties. However, ensemble classifier based on multi-feature is burdened with computational intricacy and lethargic processing. This paper presents a perspective view based license plate character recognition algorithm that copes with some of the inherent problems regarding character recognition. The proposed algorithm is resilient against slant, skewness and obscurity of image while limiting processing time and memory usage. Our algorithm tracks the principal curvature of the segmented image in terms of magnitude and inter-impulse distance. A simple back propagation neural network offers recognition rate of 93% with faster training and validation time. A thorough analysis of the algorithm and meticulous details of implementation are presented throughout this paper.**

***Index Terms*— offline character recognition, OCR, perspective view, ensemble classifier, license plate, neural network.**

INTRODUCTION

Transportation safety and mobility is a major concern in intelligent transportation system (ITS). Recent development in automatic object tracking process has been accomplished with the help of advanced technologies, which bears a wide impact in traffic monitoring as well as transportation safety. License plate recognition (LPR) has been a special area of interest in video surveillance, unattended parking lots, security control of restricted areas, automatic toll collection and traffic law enforcement [1]. As the license plate remains the principal vehicle identifier, therefore ITS rely heavily on robust LPR system.

Typically an LPR system consists of two main tasks. First one is license plate localization and the second one is identifying license numbers. Impediments regarding undesired input image conditions such as blurry images, undesired illumination conditions, small size plates, rotations, shadows and different weather conditions make generalized algorithm extremely complex. Though numerous techniques are proposed in the literature for license plate recognition system (LPR), many of them are guilty of computational complexity, sluggish training process etc.

Unlike many of the existing approaches, our study proposes a novel algorithm to recognize the license plate characters which reduces computational complexity. Proposed algorithm depends on right, left, up and down views of a segmented character. For better accuracy, top-bottom views and cross-sectional views can be taken into consideration. Later, views are filtered through a smoother and moving average filter to produce bi-directional impulses whose magnitude represent the acuteness of curvature and inter-impulse distance represent relative position of point of curvature. It has been observed that similar looking characters exhibits similar principality of curvature (POC). Four perspective views are cascaded to form a single vector set which is later used to train a simple back propagation neural network.

This approach has several inherent advantages over other feature extraction algorithm. Firstly, POC is almost independent of slant or skewness. Secondly, this method is free from the adverse effect of blurry or obscure image. Instead of matching whole image, this method tracks principal curvature of segmented characters by computing the transition of binary data. Despite the image is obscure or less vivid, our proposed algorithm performs satisfactorily. Finally, as only principal curvatures are being used for training the network instead of whole image, training and validation speed are much faster.

STATE OF THE ART

License Plate Detection

License plate detection means localization of license plate from complex background. This task is combined of edge statistics and mathematical morphology techniques where the gradient magnitude and their local variance of an image are computed [2]. High edge magnitude and high edge variance are also identified from the license plate region. This method is applicable with great accuracy in image with unclear license plate boundary within a complex background.

Color or gray-scaled based processing method is proposed in the literature [1] for license plate location. Color or Gray Scale segmentation stage is very crucial to this method. But underlying problem is that this approach becomes unstable subject to change in illumination. Besides, image may contain scattered region of similar color texture. Though the

combination between Continuous Adaptive Mean shift (CAM shift) and Support Vector machine (SVM) technique has made this method more efficient, but proposed method still could not overcome the problem of image with extremely blurry shape or complex color texture.

In order to identify the license plate precisely, various neural network architecture such as Discrete Time Cellular Neural Networks (DTCNNs), Pulse Coupled neural networks (PCNNs), Time Delay Neural Networks (TDNNs), back propagation neural network, feed forward neural network, recurrent neural network, neuro-fuzzy network and multilayer perceptron neural network are proposed and implemented in [3]. A recent observation demonstrated that computationally most intensive steps in LPR could be realized by DTCNNs . Later, a faster method which is dependent on size or distance is proposed [4]. It works by the principle of scanning a vehicle image with N row distance and counting the existent edges. With the conventional methods, parallel Fuzzy logic method has been proposed in [5]. To detect the horizontal and vertical plate position, the method provides some intuitive rules to describe the license plate and gives some membership functions for fuzzy sets. But the main drawback of this method is that it needs longer processing time and it is very sensitive to the license plate color.

Hough Transform (HT) can be used successfully where image is captured from near distance [6]. In this method, edges in the input images are detected first, and later, the license plate regions are detected by using Hough Transform (HT). But the major complexity associated with this method is that execution time of the HT requires too much computation when applied to a binary image with greater number of pixels. Based on the Adaptive Boosting (ADA boost) algorithm, a strong classifier can be trained from a set of weak classifier to identify the license plate [7]. The strong classifier is applied to sub-regions of an image being scanned for likely license plate locations. Some other popular techniques are Gabor transformation method [9] ,Wavelet transform based method [8] both of which are used to locate the multiple plates with different orientation of one image etc. are proposed in the literature.

Character Segmentation

License number identification phase consist of two major tasks: character segmentation and character recognition. Histogram based algorithm which automatically detects and merges fragments from seriously degraded images in order to segment the fragmented character is proposed in the literature [10]. To separate the overlapped character, a morphological thickening algorithm is used. On the other hand, morphological thinning algorithm determines the baseline for segmenting the connected character. In video sequence for license plate segmentation, a method is developed where the extraction of character as a (Markov random field) MRF is modeled [11]. Whole model is formulated as the problem of maximizing a posterior probability based on a given prior knowledge and observation.

More innovative method like local binarization where an image is divided in to m×n blocks was proposed previously [12]. Besides, Feature Vector Extraction and Mathematical Morphology Markov Random Fields (MRFs) Sliding Concentric Window etc. are some of the most popular techniques for segmentation of characters [2], [10], [11].

Character Recognition

Multilayer perceptron neural network can be used for recognition of characters. The network consists of a training method which uses error back propagation (BP). A more comprehensive and direct approach includes two probabilistic network, one for alphabet recognition and the other for number recognition defined as Probabilistic Neural Networks (PNNs) for LPR [13]. Recognition rates are remarkable specially for noisy, tilted and degraded pattern [13].

Pattern matching is a suitable technique for single font and fixed size character recognition if a normalized crossed correlation operator is used [11]. A normalized cross correlation value for all the shifts of each character is used in recognition process. Additionally by using four SVM , a cascaded system can be designed for character recognition [9].

Most comprehensive approach utilizes ensemble classifier which extracts diversified features of the character image. A combiner combines individual feature vectors to form a single feature class. The Output of the combiner acts as input to a neural network. Neural network is either trained with the whole image or feature class to produce suitable output.

PROPOSED ALGORITHM: PRINCIPALITY OF CURVATURE

Principality of curvature technique takes benefit of sustained curvature of a character. Firstly, it breaks down 2-D information stack into separate 1-D information set. It takes perspective view from right, left, up and down corner into consideration. For better accuracy performance, top, bottom and cross sectional view at different angles can also be considered. This technique has one major benefit is that it is independent of font size and dependent on slant, skew and obscurity of test character. The block diagram of the overall technique is provided at Figure 1.

Reading Image and Background Processing

In this paper, focus is limited to character recognition. It is assumed that background is fairly neat for segmented character image and text has been localized and separated from complex image. First step to POC is image reading and converting it into binary image. Matlab software has been used though out this paper. Figure 2 shows background processing steps and flow diagram. Next step is to impose binary threshold to obtain a binary image.

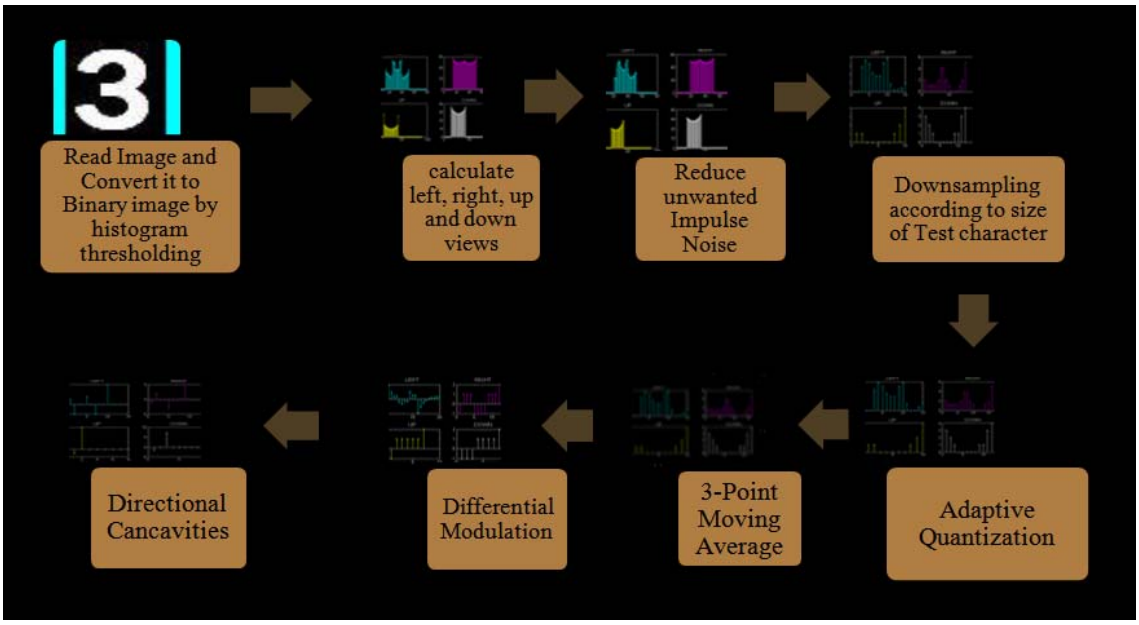


Fig. 1. Flow Diagram of Principality of Curvature

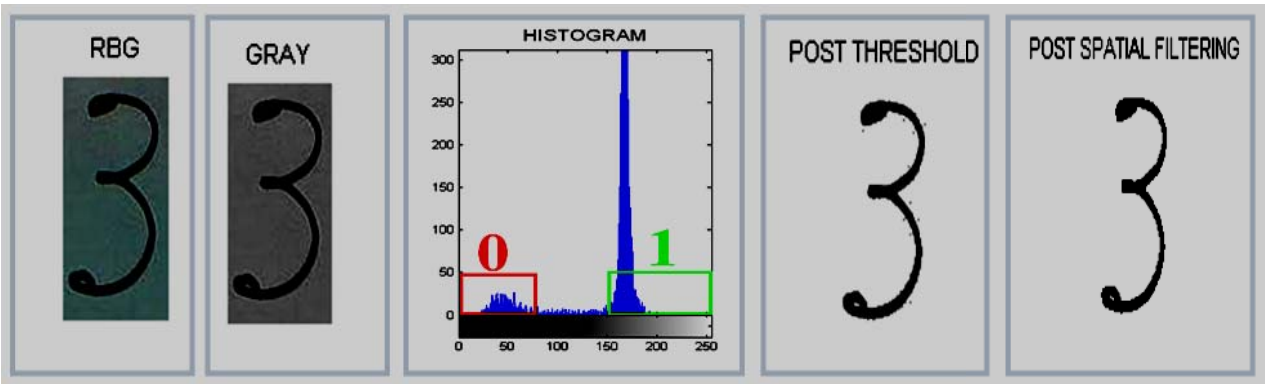


Fig. 2. Flow Diagram of Principality of Curvature

Geometric Views

Geometric views are derived as the first black pixel moving toward the opposite corner from all four sides of the image. In this paper left, right, up and down views are considered. For universal language, top, bottom and cross sectional views should also be considered.

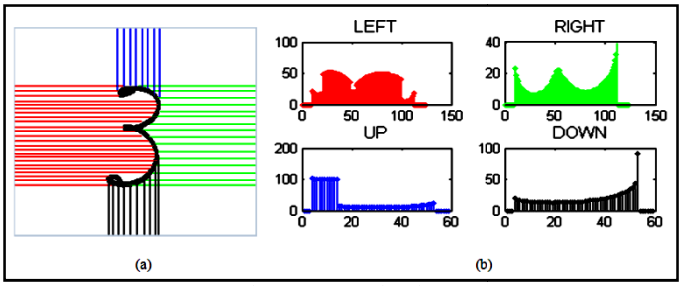


Fig. 3. Geometric Views

View Processing

View processing consists of removing unwanted noise during image conversion. It removes sudden unwanted impulse, abrupt short length directional fluctuations. This step consists of impulse reduction, downsampling according to comparison set, quantization and moving average. Downsampling is necessary to reduce computational complexity, processing time and memory usage.

Final Feature

Νοτε τηατ τηε εθυατιον ισ χεντερεδ υσινγ α χεντερ ταβ στοπ.

This step is basically divided into two parts. Firstly, it calculates successive difference between two consecutive pixels. It conserves the information of directional properties. Next step is to process this information to derive the principal curvature. Magnitude of directional impulse signifies sharpness of curvature and duration of curvature. In the meanwhile position signifies appropriate point of curvature in the test image.

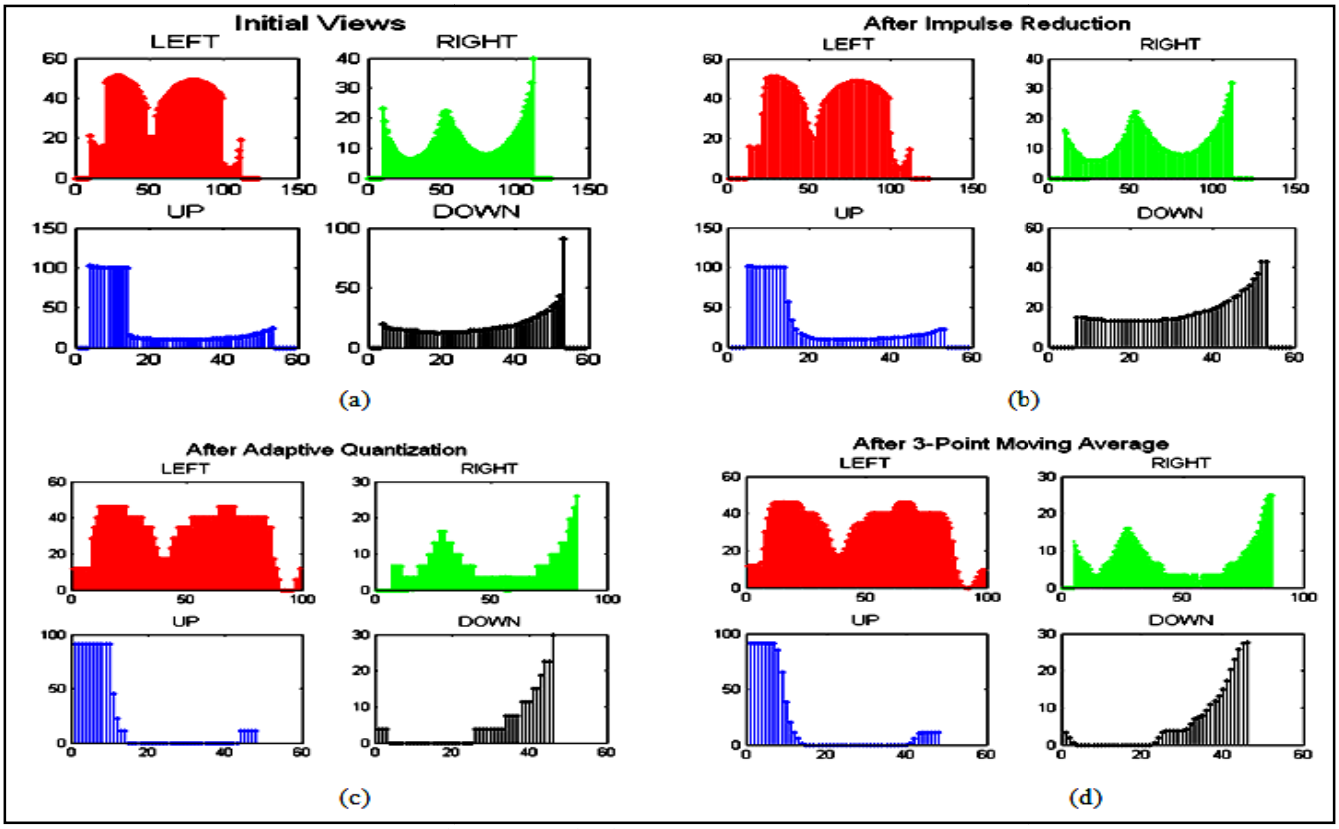


Fig. 4. Geometric Views

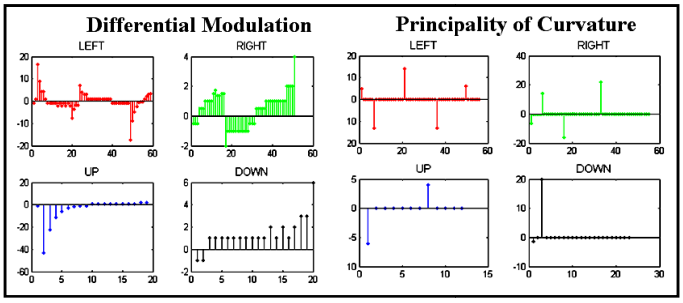


Fig. 5. Principality of Curvature

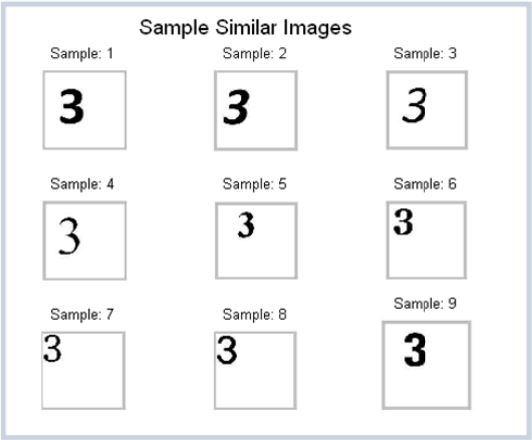


Fig. 6. Similar Looking Test Characters

It is interesting to note that Principality of Curvature (POC) is loosely dependent on font size, slant and skew. Figure 6 and Figure 8 shows features corresponding to similar looking character of different size and style.

BACK PROPAGATION NEURAL NETWORK

A combiner combines left, right, up and down views into a single feature vector which is used for training the network. A simple back propagation (BP) neural network is used with input layer of 120, hidden layer of 86 and output layer of 58.

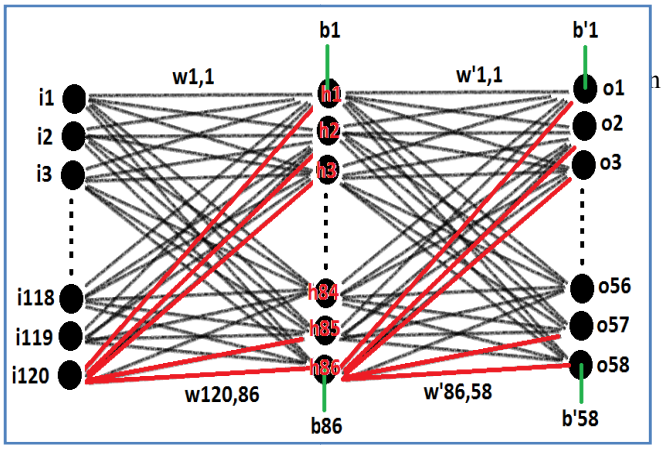


Fig. 7. BP neural network

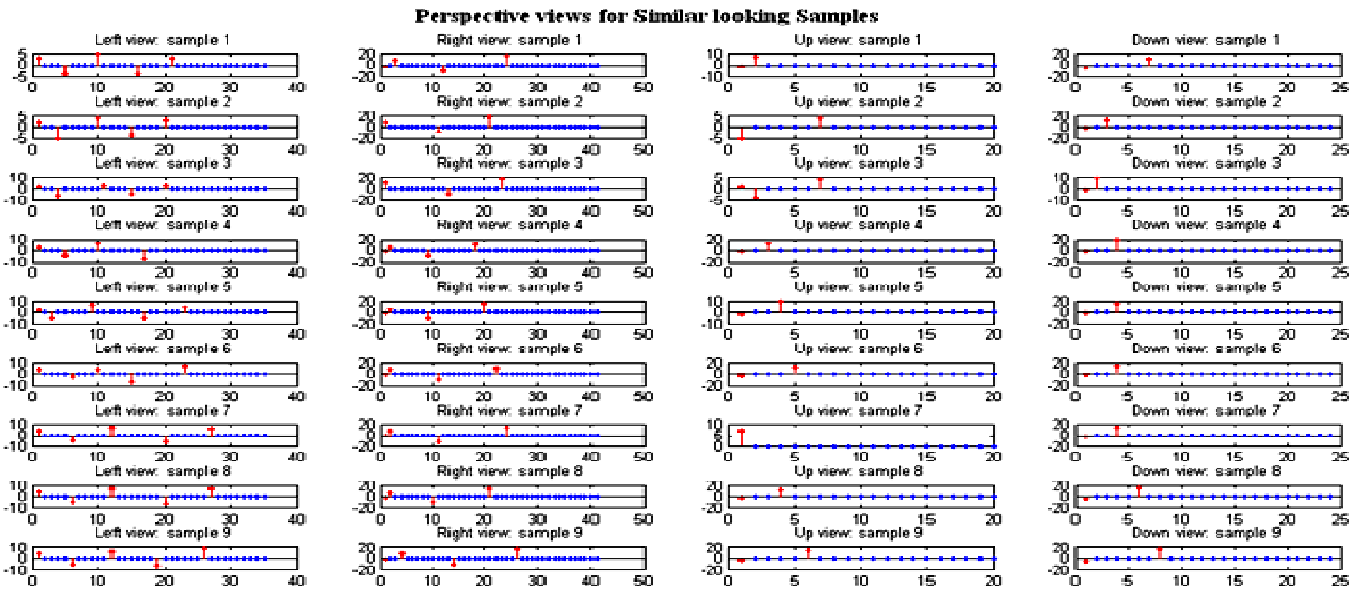


Fig. 8. POC for test characters in Figure 6.

Neural network has been successfully used for solving non-linear complex problems. It is an iterative process where network is trained by a series of update of weights based on error gradient. Initially the network is trained with standard training characters where desired result is known. Error rate is validated though number of validation characters after predefined number of iterations with the training characters. If convergence is reached and error rate is circumscribed within tolerance level, the network is ready to face any challenges. Input to the network is a vector of length of 120 comprised of left, right ,up and down views.

Starting values for weights and bias are initialized :

Desired output, [D(n)]

Input to hidden layer weight, [W(n)]

Hidden layer to output weight [W'(n)]

Hidden layer bias factor,[B(n)]

Output layer bias factor,[B'(n)]

Input and Output learning rates, γi, ζo, γo, ζi, αi, αo

Initial Weight Gradient at hidden layer, $\triangle W(n-1)$

Initial Weight Gradient at output layer, $\triangle W'(n-1)$

$$H'(n) = [W(n)][In(n)] + [B(n)] \quad (2)$$

$$H(n) = \left[\frac{1}{1 + exp^{H(n)'}}\right] \quad (3)$$

$$In(n) = \left(\sum_{n=1}^{30} I_n(left) + \sum_{n=1}^{30}(I_n(right) - 30) + \sum_{n=1}^{30}(I_n(up) - 60) + \sum_{n=1}^{30}(I_n(down) - 90)\right)^T \quad (1)$$

$$O'(n) = [W'(n)][h(n)] + [B'(n)] \quad (4)$$

$$O(n) = [\tan h\,(O'(n))] = [o1\ o2\ o3 \ldots o58]^T \quad (5)$$

$$E(n) = [D] - [O] = [d1\ d2 \ldots d58]^T - [o1\ o2 \ldots o58]^T \quad (6)$$

$$E(n) = [e1\ e2 \ldots e58]^T \quad (7)$$

$$\text{Identity Column Matrix, } I = [1\ 1 \ldots 1]^T \quad (8)$$

$$E1(n) = \left(\sqrt{(E(n))T\ E(n)}\right) * I \quad (9)$$

Next step is to compute error matrix by subtracting output matrix from desired matrix at the training stage. Error matrix is computed by Eq. 7-10.

$$O1(n) = \sum_{i=1}^{58}\left(1 - oi(n)\right) * (I - O(n))^T * [A]_{i,j} \quad (10)$$

$$O2(n) = O1^T(n) \quad (11)$$

$$\Delta O2(n) = \sum_{i=1}^{58} o2i(n) * E(n))^T * [Ai, j] \quad (12)$$

$$\Delta O(n) = (\Delta O2(n))^T \quad (13)$$

$$H1(n) = \left(\sum_{i=1}^{58}\left(oi(n)\right) * \left(I - O(n)\right)^T * [A^T]\right) \quad (14)$$

$$H2(n) = [W'(n)] * [\Delta O(n)] \quad (15)$$

$$H(n) = \left(\sum_{i=1}^{58}\left(h1i(n)\right) * (H2(n))^T * [A^T]\right) \quad (16)$$

$$W(n) = \gamma i * \left(\left(\sum_{i=1}^{58}\left(in(n)\right) * \left(\Delta\, H(n)\right)^T * [A^T]\right) \quad (17) \quad (17)$$

$$\Delta W'(n) = \gamma o * \left(\left(\sum_{i=1}^{58}\left(h(n)\right) * (\Delta\, O(n))^T * [A^T]\right) \quad (18)$$

Gradient values are used for the next iteration which will take input from another character so that previous learning

values for previous character is not lost. Eq. 19-22 are used for next iteration for next character.

$$W(n+1)=W(n)+\Delta W(n) + \zeta_i * \Delta W(n-1) \quad (19)$$

$$W'(n+1)=W'(n)+\Delta W'(n) + \zeta_o * \Delta W'(n-1) \quad (20)$$

$$B(n+1)=B(n)+\alpha i * E1(n) \quad (21)$$

$$B'(n+1)=B'(n)+\alpha o * E(n) \quad (22)$$

Training based on a complete set of character comprised of single entity from every character stack completes one iteration phase. Several iteration Phases are usually needed in order to attain desired character recognition.

V. Implementation

Open Access database MNIST and Banglabasic database were used for training, validation and testing purpose [14].

An interesting point is that generally two standard templates are followed in License plate. Either black color font is used in white background or vice versa. In case of white font, the image is inverted. Histogram is generated for each image. Generally two maxima is achieved, one for background and one for font color. If a lower maximum occurs at the vicinity of white color, the image is inverted.

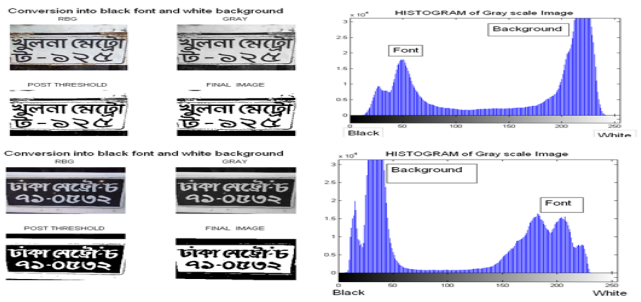


Fig. 9. Sample License Plate (a) Black font (b) Histogram (c) White font (d) Histogram

Up and down perspective views are applied on the whole license plate to separate the digits and characters. Separated digits and characters are saved as image file. Later, these extracted separated characters are used to test the performance. Fig. 10 shows the overall performance.

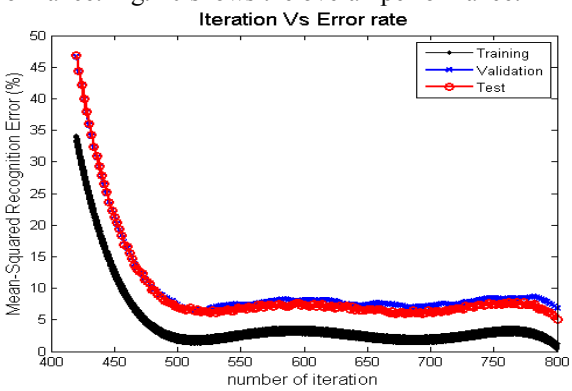


Fig. 10. Performance of proposed algorithm

V. Conclusion

This paper presents a novel feature based technique for recognition of license plate characters. Perspective view overcomes some the problems that all feature based algorithm such as font size, acute angle image capture, slant and skewness etc. Instead of whole image, only principal curvatures are used to train, validate and test the network. As a result, network has to process lesser number of information to train itself resulting a shorter training time. before the plate numbers. Numbers may contain symbols like'-'or'.'.

References

[1] *S. Draghici, "A neural network based artificial vision system for license Plate recognition," Int.J.Neural Systems, vol. 8, pp. 113–126, 1997.*

[2] Christos Nikolaos E. Anagnostopoulos, Ioannis E. Anagnostopoulos, Vassili Loumos and Eleftherios Kayafas, "A License Plate Recognition Algorithm for Intelligent Transportation System Applications," *IEEE Trans. on intelligent transportation system*, vol.7, no.3, sep-2006

[3] K. K. Kim, K. I. Kim, J. B. Kim, and H. J. Kim, "Learning-based approach, for license plate recognition," in *Proc. IEEE Signal Process. Soc. Workshop, Neural Netw. Signal Process.*, 2000, vol. 2, pp. 614–623.

[4] A. Broumandnia and M. Fathy, "Application of pattern recognition for Farsi license plate recognition," presented at the ICGST Int. Conf. Graphics, Vision and Image Processing (GVIP), Dec. 2005.[Online].Available: http://www.icgst.com/gvip/v2/P1150439001.pdf

[5] N. Zimic, J. Ficzko, M. Mraz, and J. Virant, "The fuzzy logic approach to the car numlber plate locating problem," in *Proc. IIS*, 1997, pp. 227–230.

[6] T. D. Duan, T. L. Hong Du, T. V. Phuoc, and N. V. Hoang, "Building an automatic vehicle license plate recognition system," in *Proc. Int. Conf. Comput. Sci. RIVF*, 2005, pp. 59–63.

[7] L.Dlagnekovin *License Plate Detection Using AdaBoost*. La Jolla: Comput. Sci. Eng. Dept., Univ. California San Diego, Mar.2004. [Online].Available :http://www.cse.ucsd.edu/classes/fa04/cse252c/projects/louka.pdf

[8] F. Kahraman, B. Kurt, and M. Gökmen, "License plate character segmentation based on the gabor transform and vector quantization," in *Lecture Notes on Computer Science*, vol. 2869, A. Yazici and C. Sener, Eds. New York: Springer-Verlag, 2003, pp. 381–388.

[9] S. Nomura, K. Yamanaka, O. Katai, H. Kawakami, and T. Shiose, "A novel adaptive morphological approach for degraded character image segmentation," *Pattern Recognit.*, vol. 38, no. 11, pp. 1961–1975, Nov. 2005.

[10] Y. Cui and Q. Huang, "Extracting characters of license plates from video sequences," *Mach. Vis. Appl.*, vol. 10, no. 5/6, pp. 308–320, Apr. 1998.

[11] T. Naito, T. Tsukada, K. Yamada, K. Kozuka, and S.Yamamoto, "Robust license plate recognition method for passing vehicles under outside environment," *IEEE Trans. Veh. Technol.*, vol. 49, no. 6, pp. 2309–2319, Nov. 2000.

[12] B. R. Lee, K. Park, H. Kang, H. Kim, and C. Kim, "Adaptive local binarization method for recognition of vehicle license plates," in Lecture Notes on Computer Science, vol. 3322, R. Klette and J. Uniæ, Eds. New York: Springer-Verlag, 2004, pp. 646–655.

[13] C. Anagnostopoulos,T. Alexandropoulos, S. Boutas, V. Loumos, and E. Kayafas, "A template-guided approach to vehicle surveillance and access control," in *Proc. IEEE Conf. Advanced Video and Signal Based Surveillance*, 2005, pp. 534–539.

[14] http://yann.lecun.com/exdb/mnist/

Towards Ubiquitous Learning Tools for Computer Aided Classroom in Developing Regions

Mashrura Tasnim, Farhana Zaman, Hasan Shahid Ferdous, Sharif Md. Saad Galib
Department of Computer Science and Engineering, Bangladesh University of Engineering and Technology, Dhaka, Bangladesh
mashrura_cse@yahoo.com, farhana27glory@gmail.com, hasan.shahid@csebuet.org, galib2145@gmail.com

***Abstract*— Education is recognized as one of the most powerful aspects for human development. Scope of education includes dispersal of knowledge, logical thinking, flourishing skills, refining habits, and sharing values. Accordingly, basic or primary education is a fundamental right for every citizen of a country. Studies provide sufficient evidence that individuals lacking primary education are incapable of making sensible choice about their own benefits in personal, social, and economic life. To ensure technology enhanced primary education for every citizen is yet a significant challenge for developing countries. A number of obstacles in the financial, technological, social, and political sector impede the way to exploit the blessings of modern technology in this sector. Keeping all these in mind, we have worked on a cost efficient tool to improve the learning experience in classrooms in developing regions. We have also designed some learning materials that can make learning exciting and enjoyable to the young learners. In this paper we focus on the existing problems in Primary Education system and the causes behind them. Later we present our cost effective equipment; application of which is believed to be effective in improving the scenario.**

Keywords*— *learning tool; raspberry pi; developing regions.

I. INTRODUCTION

For the last few decades the magic wand of Information Technology, with its sparkling touch, has changed every aspect of our life. It offers people all over the world an opportunity to overcome geographical distance and cost barriers to acquire knowledge. Nowadays, open, distance, and technology-facilitated learning aim to ensure availability of same resources to the learners of every corner of the world. However, challenges created by poverty, lack of social and educational infrastructures, political and religious controversies, and cultural issues restrict educational progress in developing countries including most countries in Latin America, Middle East, Africa, Southeast Asia, and some parts of southern Europe. This paper addresses some of the challenges on the way of imparting quality primary education in Bangladesh, a developing country in the South-East Asia.

Bangladesh has one of the world's largest primary education systems with approximately 16.96 million primary school aged children, 395.3 thousand teachers, and 78.6 thousand institutions [1]. Though in recent years, Bangladesh has made significant progress in promoting enrolment at primary level, progress in the quality of education in this level has been far from impressive. In our paper we have emphasized on this issue and suggested a plausible solution to the problems regarding quality of primary education in Bangladesh. The rest of the paper is structured as follows. Section II describes the related works. In Section III we have discussed the existing situation of primary education in Bangladesh. We have tried to point out the reasons behind the problems in Section IV. In Section V, we have presented our tentative solution for some of these problems. We have drawn a conclusion after discussing about Future works in Section VI.

II. RELATED WORKS

Islam [2] in his paper highlights the fact that the performance of Bangladesh has been impressive in enrolment rate and gender parity compared to other South Asian countries. However, based on several criteria, Bangladesh appears to be lagging behind other countries in the quality of primary education.

Sabates et. al. [3] identified poverty, poor health, malnutrition, lack of motivation to learn, child labor, school location and poor quality of education as major factors associated with the process of dropping out from school on the basis of survey data analysis.

On the ground of careful watch over classroom climate, teaching style, classroom management, and understanding the subject matters during the class time at ten primary schools, Shekh [4] reached the conclusion that despite various efforts, the quality of education is deteriorating and the current trend of deterioration should be dealt with, if cognitive development of the students is to be assured.

Raihan et. al. [5] emphasized on technology integration to ensure meaningful learning. In their paper, technology integration goals along with ways to use technology in real classrooms and technology teaching were discussed.

Winthrop et. al. [6] investigated the influence of technology in a variety of educational fields like: distance learning, digitized teaching and learning materials, information management and teacher support to overcome some of these existing barriers. Necessity of understanding the technological landscape in the developing world is discussed elaborately to figure out strategic uses of technology for improving education.

Whitton discussed the pros and cons of game based learning in her paper [7]. Here the option of developing simple

and cost-effective games with low technical specifications is explored. Feasibility of learning from games rather than with them is analyzed. Finally, the option of giving learners agency as game creators rather than simply players to facilitate learning is presented.

The study conducted by Ang et. al. [8] showed that students get more interest and engagement in completing the learning task when lessons are presented as 3D game. According to their observation, participants expected that, more scientific topics would be presented in such a way. In conclusion, some issues regarding student engagement and implications for instructional design were discussed.

From the above analysis it is evident that to achieve the goal of ensuring quality education, cost-effective technology has to be adapted by developing countries. In our work we will present our work-plan in accordance with this fact.

III. Present Situation of Primary Education in Bangladesh

Primary education has been a priority in Bangladeshi policy making since independence, basic measures to implement universal primary education were taken from the outset. At present Bangladesh has a five years primary education system. Children of the age 6 to 10 are offered primary education which is compulsory by law and free of costs. There are several types of schools offering primary education throughout the country. As a result Bangladesh has achieved significant progress in enrolment in primary schools. Net Enrolment Rate (NER) was 93.52% in 2010 [9]. However, according to the Global Competitiveness Report, 2011-12, among 142 countries, Bangladesh ranks 113th in primary enrolment and 119th in the quality of primary education.

Disparities in the quality of education are by far the biggest challenges Bangladesh faces in the primary education sector. Currently, two major problems of this sector are: dropout before completing primary education and poor quality of education with low attainment of basic competencies.

A. *Dropout*

Although entry into Grade 1 has reached near universal level, the primary school completion rate is lagging far behind. Annual Education Survey (AES) 2010 [10] shows the alarming drop-out rate of 45.1% in primary level; 47.8% and 42.5% for boys and girls respectively. According to the statistics of Directorate of Primary Education (DPE), around 3.79 million students of different types of schools registered for enrolling class-I in 2008 [12] and of them 2.64 million registered for Primary School Certificate (PSC) Examination in 2012. As this examination is treated as primary completion examinations of general and madrasa education in Bangladesh, this statistics shows around 30% students were dropped out from their academic life in different grades from I to V during the period of 2008-2012.

B. *Quality of Education*

Quality of education is a multi-dimensional term and determined by objectives of education, therefore is difficult to measure with quantifiable indicators. According to EFA declaration, 2000 by The World Education Forum; quality of education is defined as "Ensuring excellence of all so that recognized and measurable learning outcomes are achieved by all, especially in literacy, numeracy and essential life skills" One approach of measuring the quality of education is in terms of student performance in standardized tests in basic language, mathematics and science. Although in 2012 Primary School Certificate Examination pass rate reached 97.35%, some other studies reveal an opposite scenario. Bangladesh Literacy Survey 2010 [11], covering 43,945 households (34,620 from rural areas and 9,325 from urban areas), treated capability of writing a letter for communication as parameter of quality. The findings of this survey indicate that only 46.47% of the participants succeeded to write the letter after receiving 5 years' primary education.

Besides socio-economic factors such as poor health, malnutrition, lack of motivation to learn, child labor and poverty, school level factors such as teacher absenteeism, school location, infrastructure and poor quality learning materials are also identified as influencing reasons for above mentioned problems [3]. In section IV we have discussed some of these reasons in detail.

IV. Reasons Behind the Scenario

In the 'Report on Primary Education in Bangladesh: Challenges and Successes' produced by Bangladesh Bureau of Education Information and Statistics (BANBEIS) and Bangladesh Ministry of Women and Children Affairs (MWCA), a number of reasons responsible for the low quality of primary education were pointed out [13]. Some of the issues are as follows:

A. *Insufficient Spending on Education*

An amount of Tk. 214080 million was allocated in the budget for Fiscal Year 2012-13 for the Ministry of Education and Ministry of Primary and Mass education to conduct both development and non-development activities. This allocation is 11.17 percent of the national budget of Fiscal Year 2012-13 [14]. However, 96% of the revenue for education in government schools is spent to provide teachers' salaries leaving little money to train and supervise new teachers and to improve classroom environment.

B. *Poor School Facilities*

Building and improvement of new facilities are two of the main focuses for governmental education. Yet most schools are far from having adequate facilities. Classrooms are not clean or safe; they lack basic necessities such as textbooks, blackboards, electricity etc.

C. *Difficulties in Access to Education*

Government schools are concentrated mostly in urban areas, resulting in a lack of educational access to the majority

of children in the country. Even considering the high number of people moving to cities, the majority of people in Bangladesh live in rural areas. Many of the government schools are too far away or too crowded, discouraging parents from sending their children to school. Girls, children from rural areas and urban periphery areas, children from socioeconomically backwards families and ethnic minorities are often denied access to the governmental education system.

D. Class Size and Attendance

Most primary schools do not prioritize small classes; often student to teacher ratios are as high as 60:1. Classrooms are unable to properly accommodate all students. Overcrowded classrooms and lack of facilities are contributing factors to the 60% attendance rate of children being able to attend school.

E. Contact hours

Due to lack of classrooms, majority of primary schools are operated on a double-shift basis, resulting in annual contact hours between teachers and students of only 595 for Grades 1 and 2 and 865 for Grades 3 to 5, resulting in serious negative effects on the quantity and quality of learning [15].

F. Teacher Training and Supervision

In order to teach at a government school in Bangladesh, teachers have to attend a yearlong training at a Primary Teacher Institute (PTI). This training period is meant to assure a high standard of teacher competency. However, teachers often do not use most of what they learn and there is little supervision and few refresher training sessions after the initial year training. Teachers in government schools have been found to be passive and interact little with students.

G. Curriculum

Rote-learning is the main style of curriculum. Textbooks, being irrelevant to real life, often fail to interest students. Few supplemental materials are provided. Most schools also lack extracurricular activities. Students are given large amounts of homework to be completed at home, and parents are illiterate or uneducated to help their children. Children from poor families who do not have the capability to hire tutors often fail to complete their homework which leads to the children's and the teachers' frustration resulting in higher failure and dropout rates.

H. Unfriendly learning environment

Lack of entertainment facilities is viewed as a potential reason of dropout by many teachers and local government representatives. About 8% parents of dropped out children also identified lack of entertainment facility for students as a cause of dropout. According to them there is no arrangement for any kind of entertainment which would attract the students towards schools. Some teachers also pointed the lack of entertainment facility makes education unfriendly. They also think that there is also a lack of learner friendly and need based curriculum [16].

V. A Tentative Solution

In this section we present our findings that can be implemented to improve the quality of Primary Education. It consists of two parts. We have discussed them in detail here.

A. Raspberry PI

The Raspberry PI is anticipated as a twinkling opportunity for school kids of developing countries to get in touch of computer based education. It is a credit card-sized low cost computer. Since its launch in February 2012, it has been sold an impressive 1.2 million around the world.

The ARM-based PC features a HMDI socket, a video output, USB ports, Ethernet, audio output, and SD card slot. Power is provided either by a battery or a standard mobile phone charger. Model B comes up with 512 MB of RAM and two USB ports at $35; while Model A, providing 256 MB RAM and one USB port, costs $25 [17].

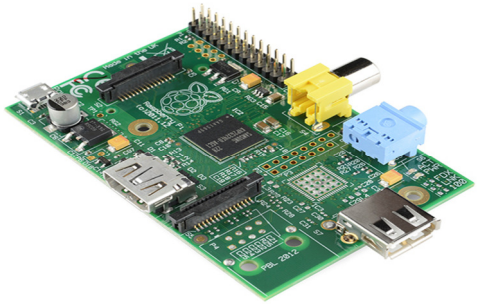


Fig. 1. Raspberry PI Model A.

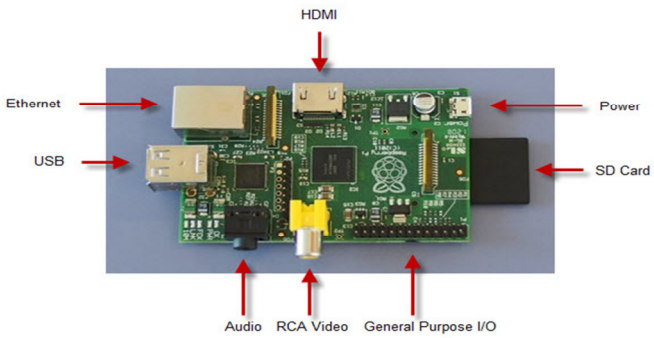


Fig. 2. Raspberry PI Model A.

Raspberry PI is a fully programmable PC, runs a free, open-source Linux operating system, plugs into any TV, can power 3D graphics, and connects to the Internet. The tiny computer has been proved a fun project for small computing tasks. It can bring about a revolutionary change in education system of developing countries like Bangladesh.

By introducing use of Raspberry PI in a wide range, many of the problems discussed earlier can be reduced or even eliminated. It will enable students to access the vast resources

available on the web and thus dependency on private tutor can be reduced. Several organizations, such as Khan Academy, produce short educational videos that serve as self-learning tools, curricular supplements or the core teaching materials in classrooms where teachers may not be experienced enough in particular subjects. Besides, by taking help of Google translation and transliteration, students will be able to access unlimited open educational resources from web. Getting acquainted with the tremendous variation of resources will flourish their creativity, widen their insight and strengthen their logical and analytic capacity.

In this regard it can be mentioned that a similar project of assembling laptops locally was launched by Bangladesh Government in collaboration with state-owned telephone manufacturing company TSS, 2M Corporation Ltd and TFT Display (M) Co of Malaysia, aiming to offer the ICT device at affordable prices to the local users, especially students. The price range was declared ranging from 130$ to 365$. However this project named 'Doel' laptop could not celebrate a big success mostly due to quality issues and higher price than that declared earlier [19]. On the other hand Raspberry PI offers much lower price and response from user end bears testimony of quality service of this device.

Education works at its best when delivered in mother tongue, especially in primary level. So besides introducing low-cost computers, it is also important to provide appropriate online lessons for curious learners. Recently Bangladesh Rural Advancement Committee (BRAC) has introduced Computer Aided Learning (CAL) for students from grade 6 to 10 with a view to developing interactive education software [18]. We have taken a similar initiative to design and develop attractive online learning tools for primary level students, a brief description of which is given below.

B. *Learn In Fun*

Several studies show that graphics aided and digital game-based approaches present an excellent opportunity to enhance learning capacity of the students and produce a wide range of educational benefits in comparison with conventional classroom instruction. Being motivated by this finding, we have initiated the 'Learn in Fun' project, the goal of which is to represent the lessons included in the curriculum in an easy and enjoyable way. Currently most of the topics covered in grade I Mathematics are available online [20]. A few modules of the project are presented below

1) Introducing Numbers: This module is prepared to introduce numbers from 1 to 30 with attractive graphical objects. It also imparts primary know-how of counting. Fig. 3 and Fig. 4 demonstrate two examples, showing the way of introducing numbers from 1 to 10 and from 11 to 30 respectively.

2) Odd-even: It teaches the students the difference between odd and even numbers on the basis of making pairs of various well-known objects. Example of this module is presented in Fig. 5 and 6.

3) Addition: To get the basic idea of addition, users have to choose two inputs from the input panels. Then the simulation of the addition is shown step by step. It is especially helpful to learn how to handle carry. Fig. 7 demonstrates this module.

Fig. 3. Introducing numbers from 1 to 10.

Fig. 4. Introducing numbers from 11 to 30.

Fig. 5. Odd numbers.

Fig. 6. Even numbers.

4) Subtraction: Minuend and subtrahend are chosen by the user, ranging from 1 to 50. First minuend number of fishes appears in a pool, then an animated shark eats up subtrahend number of fishes and the final result is displayed. An example is presented in Fig. 8.

Fig 7. Addition.

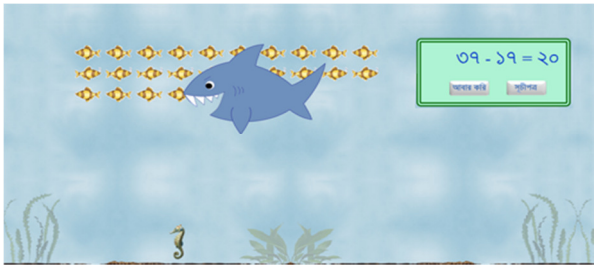


Fig. 8. Subtraction.

5) Comparison: Comparison between two inputs is simulated in this module. Difference among three results of comparison i.e. whether one of the user provided two inputs are equal, greater than or less than the other is explained in a convenient way through attractive graphics, example of which is given in Fig. 9.

Fig. 9. Comparison.

6) Ordering: This module explains basic concept of ordering of numbers through animation. Example of this module is provided in Fig. 10.

VI. Discussion and Future Works

Our initial experience of working with Raspberry PI is very much convincing and we believe it as a golden opportunity to make the blessings of ICT available to young learners of Bangladesh. The 'Learn in Fun' project has received quite impressive feedback at initial field tests. Kids were excited to learn in this new environment. Their parents and teachers also appreciated our effort. We have plans to provide a lot more lessons and an option of interactive self-test is already under development. Therefore for primary school kids a lot more excitements are yet to come.

Fig. 10. Ordering.

VII. Conclision

Bangladesh is one of the most densely populated countries in the world; therefore developing human resource is the only way for us to keep pace with the advanced world. Importance of primary education can in no way be denied in this regard. Raspberry Pi is a like spark of hope, the usability of which is yet a lot to be explored. Learn in Fun is the first website in Bangladesh developed in Bangla following the national curriculum. Web materials like this can go a long way to make the classroom environment more enjoyable. In our work we focused on these facts in a hope of drawing attention of application designers and concerned policy-makers.

References

[1] BANBEIS - Bangladesh Bureau of Educational Information and Statistics,http://www.banbeis.gov.bd/webnew/index.php?option=com content& view=article&id=119&Itemid=1

[2] Islam, Sadequl. "Access with quality in primary education: Re-inventing inter-organisational synergy." Bangladesh Education Journal 9.1 (2010): 5-27.

[3] Sabates, Ricardo, Md Altaf Hossain, Keith Lewin. "School drop out in bangladesh: New insights from longitudinal evidence". Consortium for Research on Educational Access, Transitions and Equity, 2010.

[4] Shekh, M. Nurul I. "The State of Primary Educational Quality in Bangladesh: An Evaluation". The Third NETREED Conference, Beitostolen, Norway, 2005.

[5] Raihan, Md Abu, and Seung Lock Han. "Technology Integration for Meaningful Learning-the Constructivist View". Manzoor Ahmed, 2012.

[6] Winthrop, R., Marshall S. Smith. "A NEW FACE OF EDUCATION BRINGING TECHNOLOGY INTO THE CLASSROOM IN THE DEVELOPING WORLD". Global Economy and Development, Brookings, 2012.

[7] Whitton, Nicola. "The place of game-based learning in an age of austerity". Electronic Journal of e-Learning, 10.2, 2012.

[8] Ang, Kim Hock, and Qiyun Wang. "A case study of engaging primary school students in learning science by using Active Worlds." Proceedings of the first international LAMS conference. 2006.

[9] Ardt, K., Hastings, C., Hopkins, K., Knebel, R., Loh, J., Woods, R. "Report on primary education in bangladesh: Challenges and successes". Bangladesh Bureau of Education Information and Statistics (BANBEIS)/ Bangladesh Ministry of Women and Children Affairs (MWCA), 2005.

[10] Global Competitiveness Report, 2011-12. http://www3.weforum.org/docs/GCR2011-12/16.GCR2011 2012DTIVHealthandPrimaryEducation.pdf

[11] "Report on the Bangladesh Literacy Survey, 2010". Bangladesh Bureau of Statistics. http://www.bbs.gov.bd/webtestapplication/userfiles/image/Survey%20reports/Bangladesh%20Literacy%20Surver%202010f.pdf

[12] Directorate of Primary Education (DPE) http://www.dpe.gov.bd/index.php

[13] Ardt, Kalene, et al. "Report on primary education in bangladesh: Challenges and successes." Bangladesh Bureau of Education Information and Statistics (BANBEIS)/Bangladesh Ministry of Women and Children Affairs (MWCA) (2005).

[14] Onneshan, Unnayan. "Faltering Growth: A Rapid Assessment of National Budget 2012-13', Bangladesh Economic Update, vol. 3, no. 5, June 2012." (2012).

[15] Nath, Samir R., and AM Raza Chowdhury. State of primary education in Bangladesh: Progress made, challenges remained. Campaign for Popular Education, Bangladesh, 2009.

[16] Norul Alam and Mizanur Rahman. "Primary education: Who drop out and why?" http://www.thefinancialexpress-bd.com/index.php?ref=MjBfMTJfMDJfMTJfMF8zMV8xNTE3MTk=&feature=c3BlY2lhbHNOZXdz&na=RkUgRURVQ0FUSU9O

[17] Raspberri PI, http://www.raspberrypi.org/archives/2180

[18] BRAC Education Program. http://e-education.brac.net/

[19] Jamal Uddin . "Fund crisis, technical glitches put 'Doel' laptop project in limbo".http://www.thefinancialexpress-bd.com/index.php?ref=MjBfMTFfMTVfMTJfMV8xXzE1MDAyMg

[20] Learn In Fun. http://buet.ac.bd/cse/mashrura/index.html

Interplanetary Network: A brief introduction

K. M. Sabidur Rahman
Depertment of Computer Science
University of Texas at San Antonio
San Antonio, USA
Email: wmr727@my.utsa.edu

Md. Mainul Islam
Depertment of Computer Science
and Engineering
Bangladesh University of Engineering
and Technology
Dhaka, Bangladesh
Email: mainul98@gmail.com

Md. Humayun Kabir
Depertment of Computer Science
and Engineering
Bangladesh University of Engineering
and Technology
Dhaka, Bangladesh
Email: mhkabir@cse.buet.ac.bd

***Abstract*—Interplanetary Network defines the architecture and protocols necessary to permit inter-operation of the Internet residents or systems on earth/spacecrafts with other remotely located Internet resident/systems on other systems/spacecrafts in transit in the hostile and unpredictable environment of space. Inter-networking in such environment requires new techniques other than traditional communication protocols. This paper focuses on the understanding of the current system and protocols used in Interplanetary communication. This paper is mainly based on the study of different layers of CCSDS (Consultative Committee for Space Data Systems) protocol, which is the recommended protocol used by all the Space Communication Organizations including NASA. The resources used for this paper are mainly reports from CCSDS, NASA and JPL. This paper will pave the easier way of studying and understanding of Interplanetary Network. In the CCSDS protocol Stack, Security is shown as a protocol layer. But, the reality is, every layer has its own security concerns. In this paper, we have tried to reflect this idea too.**

Keywords*—*Interplanetary Network; security; protocol;

I. Introduction

The objective of the Interplanetary Network is to define the architecture and protocols necessary to permit inter-operation of the Internet residents on Earth with other remotely located Internet residents on other planets or spacecrafts in transit. In such a hostile and unpredictable environment traditional networking protocols, such as TCP/IP, do not perform well. For this reason, CCSDS has proposed new protocol layers employing new technique for space networking. In this paper, we describe CCSDS protocol layers in brief. Our goal is to summarize the protocols in one paper so that anyone can have the overview of the whole system by just looking into one document. Spacecrafts and Space Shuttles from all over the world are using these protocols [1]. Currently NASA investigate on DTN with the goal of advancing the maturity and heritage of the DTN communication protocols in Space Communication. The DTN protocol suite is still under active development. In the rest of the paper, we shall go through the protocol layer by layer in a top-down approach standardized by CCSDS [2] and some features related to DTN.

II. Protocol Stack

The CCSDS protocol stack is very similar to the OSI Model Protocol Architecture. But, the underlying supporting protocols and mechanisms are way complex and robust to support the hostile and unpredictable environment of space. Thus, many features and attributes of Delay Tolerant Network(DTN) are present in this protocol architecture. Most of these standard layers contain multiple options for protocols that can be selected to meet mission needs. Working from the top to bottom; the layers are as follows.

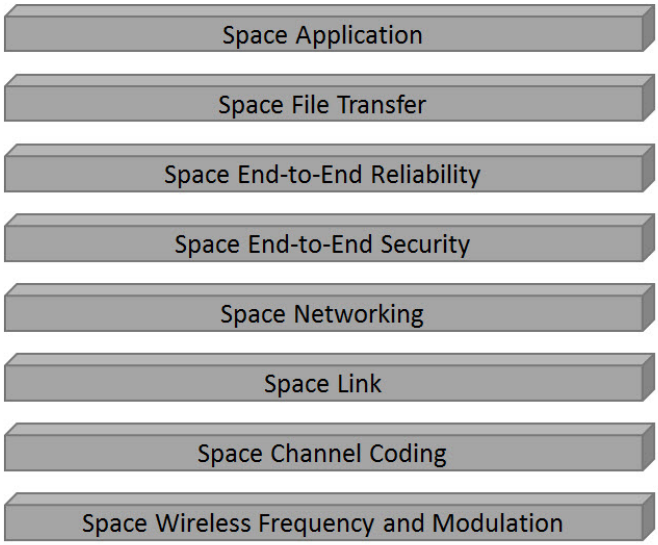


Fig. 1. Reference Protocol Stack [2]

- **Space Application Layer:** Space communications protocols of the Application Layer provide end-to-end application services such as file transfer and data compression.
- **Space File Transfer Layer:** The CCSDS currently supports two file- based standard capabilities in their File Transfer Layer:
  1) The Internet File Transfer Protocol (FTP), and SCPS space-adapted extensions to FTP. These are primarily for use in short-delay Internet-like environments, and assume an underlying layer of TCP.
  2) The CCSDS File Delivery Protocol (CFDP). This is a delay tolerant protocol whose model of operations is fundamentally store-and-forward, much like e-mail that conveys

files as attachments. The protocol as currently designed contains its own reliability mechanisms and does not assume an underlying retransmission capability.

- **Space End-to-End Reliability Layer:** The End-to-End Reliability of the protocols are ensure in this layer. In many practical implementations, the functionality of this layer is ensured by other layers and protocols such as SCPS Transport Protocol (SCPS-TP) in Transport Layer.
- **Space End-to-End Security Layer:** As missions become more Internet-accessible, they become more vulnerable to attack. Basic authentication and encryption can be accomplished within the CCSDS Link standards but more powerful end-to-end techniques can protect the entire flow of user data. Two standard protocol choices exist: Internet Protocol Security (IPSec) and a SCPS Security Protocol (SP). Both provide multiple levels of data protection.
- **Space Networking Layer:** The Space Networking Layer standards both satic and dynamic routing mechanism which supports fully connectionless communication. The CCSDS Packet has been in use as a "CCSDS Path" (connection oriented) networking protocol for well over a decade. It exploits the fact that for most current missions there is a highly predictable data routing path between an instrument and a user, so there is little need for adaptive packet routing.
- **Space Link Layer:** Space Link Layer specifies methods of transferring data units provided by the higher layer over a space link using data units known as Transfer Frames.
- **Space Channel Coding Layer:** This layer specifies methods of synchronization and channel coding for transferring Transfer Frames over a space link. Often this layer is implemented along with the Space Link Layer.
- **Space Wireless Frequency and Modulation Layer:** These standards specify the frequencies and efficient modulation types to be used to create the channel connecting the spacecraft to other spacecraft.

Although, the CCSDS standardize a list of layer, the actual implementations uses a Protocol Stack similar to TCP-IP Protocol. In the following sections we will go through the layers one by one in more details.

III. CCSDS File Delivery Protocol (CFDP)

The *CCSDS File Delivery Protocol(CFDP)* [3] provides End-to-End Application-Layer Services which is based on Internet File Transfer Protocol (FTP) and is generally interoperable with it. It extend and modify FTP in order to meet the unique requirements of space flight operations.

A. Architecture

The CFDP is designed to provide the features and attributes of Delay Tolerant Network(DTN) in Transport Layer. This protocol consists of five architectural elements:

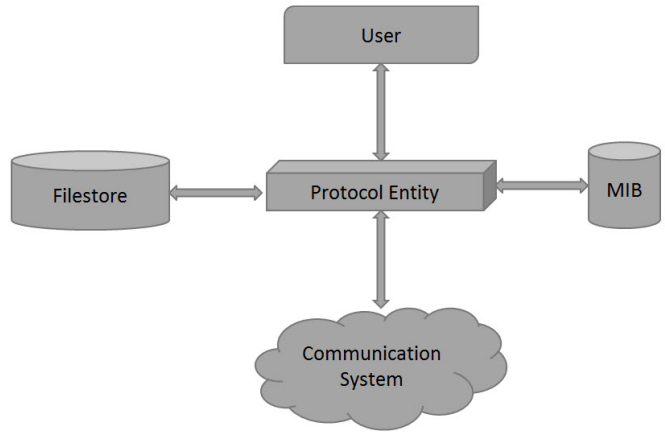


Fig. 2. Architectural Elements of CFDP

- **Protocol Entity:** The protocol entity implements the Core delivery procedures, which allow immediate file delivery and manipulation over a single network hop. Sometimes, the *Extended* procedures are available for file transfer, which allow for time-disjunct or immediate delivery over a number of network hops with appropriate facilities for onward routing.
- **User** There must a CFDP user who will request the protocol to operate. A CFDP user is always a software task, which may or may not be operated by a human.
- **File Store:** The protocol operates by copying files from one storage medium to another storage medium. For this reason, it is assumed that all CFDP entities have access to a local storage capability. As the storage capability provided may vary, the protocol is built on the premise that any file or organized set of files (i.e., a filestore) can be described in terms of a single standard representation. This representation is called a *virtual filestore*. *Virtual filestore* is assigned a standard set of attributes which are then used by the protocol to manage the file delivery process.
- **Communication System:** The protocol assumes the availability of a single conceptual underlying communication system, referred to as the *Unitdata Transfer (UT) Layer*, to which all CFDP entities in a given CFDP addressing domain have access. *UT addresses*, to which CFDP entities names can be mapped using information contained in the management information base (MIB).
- **Management Information Base:** A significant amount of information must be passed by the local user to its local CFDP entity and by the local CFDP entity to the remote CFDP entity to perform a file delivery. Usually, this data is static and may be stored within the CFDP entities as system tables, referred to as the Management Information Base (MIB). The MIB contains such information as default values for user communications requirements, e.g., for address mapping and for communication timer settings.

B. Features and Services

The CFDP support the features that allow to create and exchange files, manage remote file systems and automatically retransmission of missing or damaged piece of files in single point-to-point link. The following subsections covers the features supported by CFDP. More details of this operations are available at [3].

1) **File Copy:** The initiation of the transaction causes a *file copy* operation to occur between two CFDP entities, the sender and the receiver. All copy operations start with transmission of the Metadata, followed by transmission of the file segments and a PDU to indicate the End of File (EOF). Optional service primitives may be issued at the receiving entity to announce receipt of individual file segment. However, notification of receipt of Metadata is mandatory.

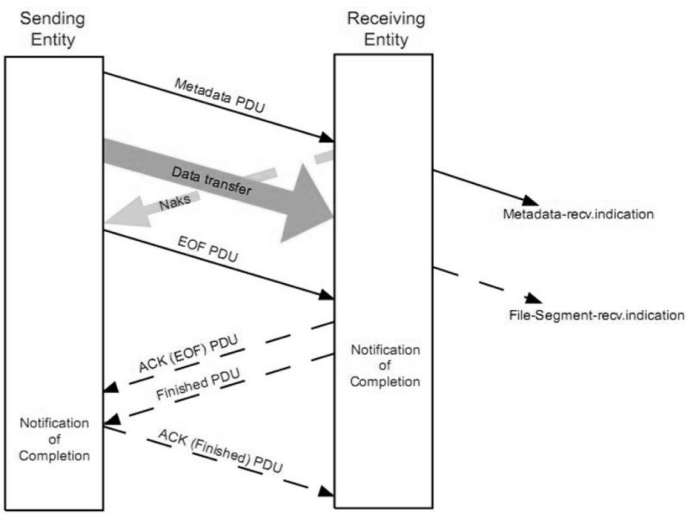


Fig. 3. Sequence of Events in Copy Operation

For reliable transactions, a number of features are added. NAKs are used to request retransmission of lost data. Receipt of the *EOF PDU* is ensured via an EOF acknowledgment. A *Finished PDU* is sent when all file data has been successfully assembled, because lost data may still be out on the way even after the EOF sequence is received. Delivery of this PDU is ensured through use of the *Finished acknowledgment*.

2) **Addressing:** Within any single CFDP network (i.e., any set of CFDP entities that is closed under same communication), each CFDP entity is assigned a unique name. Each CFDP entity has access to exactly one filestore. Identification of a CFDP entity implicitly identifies the associated filestore. Each CFDP entity can look-up for an address using information contained in the associated **MIB**. This look-up capability provides translation between the network-unique name of a CFDP entity and the corresponding UT address, which may be an Internet address or radio device buffer, APID, virtual channel number or other implementation-specific mechanism. The names or *entity IDs*, that uniquely identify the source CFDP entity and the destination CFDP entity must be included in the header of each PDU in order for an FDU to be successfully transmitted through a series of waypoint CFDP entities.

C. Security Issues and Measures taken

Application layer data and system security is ensured by the use of traditional Confidentiality, Authenticity and Integrity ensuring mechanisms. The main security issues currently considered are: Modification of messages and compromise of data confidentiality. To prevent this attacks traditional measures like Advanced Encryption Standard (AES) (to ensure confidentiality), digital signature(to ensure authentication) and Secure Hash Algorithms (SHA-1, SHA-2) (to ensure Integrity) are used. But, more robust and secure design is needed in each layer to ensure protection against all sorts of attacks. Moreover, privacy and trust issues are still not being considered in peer to peer networks for cluster of spacecrafts.

IV. Space Networking Layer

Space networking layer support communication environments with static, highly managed communication using fully connectionless communication with dynamic routing. It also supports for precedence (priority) based handling and multiple routing options. CCSDS has developed two protocols for the Network Layer:

a) Space Packet Protocol

b) SCPS Network Protocol

These protocols are developed to transfer data (1) from a source on a spacecraft to one or multiple destinations on the ground or on (an)other spacecraft, or (2) from a source on the ground to one or multiple destinations on one or multiple spacecraft.

A. Space Packet Protocol

The Space Packet Protocol provides the capability to transfer processed data efficiently over space links and the capability for routing data through a space data system [4]. Space missions will use this protocol to transfer space application data over a space-to-space communications link network. The Space Packet Protocol provides a unidirectional data transfer service from a single source user application to one or more destination user applications through one or more subnetworks. The path from the source user application to the destination user application(s) through the subnetwork(s) is called a Logical Data Path (LDP).

As the data traverse the subnetworks of the LDP, they are carried by subnetwork-specific mechanisms using protocols provided by the subnetworks. The selection of protocols used in the subnetworks is determined independently for each subnetwork and may not be the same through the entire LDP. Every LDP is uniquely identified by a Path Identifier ID. Figure 4 shows an example of an LDP from a single source user application to a single destination user application.

B. SCPS Network Protocol

SCPS-NP provides a variety of addressing and routing capabilities, some of which are compatible with the Internet, and can be used with SCPS or Internet upper-layer protocols [5].The protocol data units of SCPS-NP are called SCPS-NP datagrams. The SCPS Network Protocol (SCPS-NP) uses

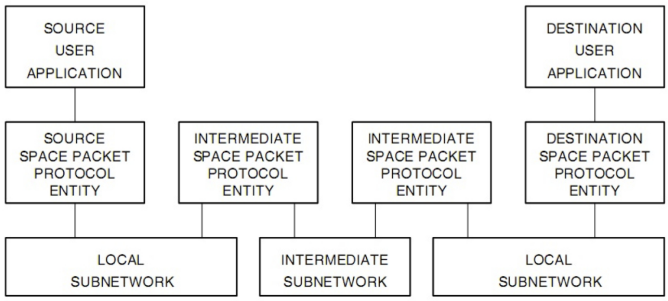


Fig. 4. Example of a Logical Data Path

a technique called capability-driven header construction as a means to control bit overhead. Capability-driven header construction simply means that the format of the SCPS-NP header is based (exclusively) on the protocol capabilities required for the communication of the particular datagram in question. That is, a datagram carries those header elements that are required to provide service properly to the datagram, but no others.

The capabilities required to support a datagram are derived from three sources: the protocol itself, the operating environment, and the network service user. Some capabilities are required to support a particular protocol version. An example of this is the capability of datagram length delimiting. Other capabilities are required to address particular network environmental conditions. An example of an environmentally required capability is header error protection. The third source of capability requirements is the network service user. An example of a user-required capability is precedence (priority).

The capability requirements from these three sources are used to select the set of header elements necessary to provide the required services in the transmission of the datagram. One key header element is a variable-length control field that identifies the remaining header elements that are included in the datagram. This control field is present in all versions of the SCPS-NP, and is the last header element before those header elements specified in the control field.In addition to user data transfer, this protocol specifies the means for exchanging network-layer control information through the SCPS Network, and for selectively passing that information to SCPS Network users. Protocol details like frame format and addressing are considered out of the scope of the paper.

C. Security Issues and Measures taken

Security issues of this layer includes flooding, broadcast storm and replay attacks. The current designs are not vary much aware of this kinds of attacks. Only a full scale modeling and implementation of Network layer Security can ensure the expected level of safety from such attacks.

V. Space Data Link Layer

A. The Space Data Link Protocols

If there is no end-to-end path between source and destination, called *network partitioning end-to-end communication*, using the existing protocols does not work. In the space network, it is not possible to establish and maintain a continuous communication path between local and remote endpoints. There is no assurance that one of the *endpoints* of a communication will be on Earth. Even if a relatively continuous communication system could be developed to support communication between Earth and Mars, for example, other elements of the networks would not be so richly connected.
The Space Data Link Protocols belong to the Data Link Protocol Layer, which specifies methods of transferring data supplied by the higher layer over a space link using data units known as *Transfer Frames*. Above the Space Data Link Protocols, higher-layer protocols belonging to the Network Layer and above use the services provided by the Space Data Link Protocols. Most of these higher-layer protocols are responsible for transferring data from end to end across the entire network, which contains the space links. Space data link protocols are similar to the Data Link Layer of the OSI Basic Reference Model, which are designed by CCSDS to meet the requirements of space missions for efficient transfer of space application data of various types and characteristics depending on space links. CCSDS has developed four Space Data Link Protocols:

- The TC Space Data Link Protocol [6]
- The TM Space Data Link Protocol [7]
- The AOS Space Data Link Protocol [8] and
- The Data Link Layer of the Proximity-1 Space Link Protocol [12]

These protocols are used to maintain communications link between two spacecraft. This consists of one forward Physical Link for sending commands from one spacecraft to the other spacecraft (or from the controlling spacecraft to the target spacecraft) and one or multiple return Physical Links (possibly using multiple frequency bands) for sending telemetry from the spacecraft to the other spacecraft (or from the target spacecraft to the controlling spacecraft). It is possible to use one instance of the Space Data Link Protocol on a Physical Channel in one direction and another instance on another Physical Channel in the other direction.

- The *TC-SDLP* is usually used for sending commands from a spacecraft to another spacecraft) on a forward Physical Link.
- The *TM-SDLP* is usually used for sending telemetry from a spacecraft to another spacecraft on a return Physical Link.
- The *AOS-SDLP* is also used for sending telemetry from a spacecraft to another spacecraft on a return Physical Link, but it may be used on a forward Physical Link as well if there is a need for two-way on-line communications (e.g., audio and video) between a spacecraft and a ground system (or between two spacecraft).
- The *Proximity-1 Space Link Protocol* is a two layer 'Space Link Protocol' specially designed for maintaining the communications among probes, landers, rovers, orbiting constellations and orbiting relays. When commands are sent from Earth, *Relay Command*

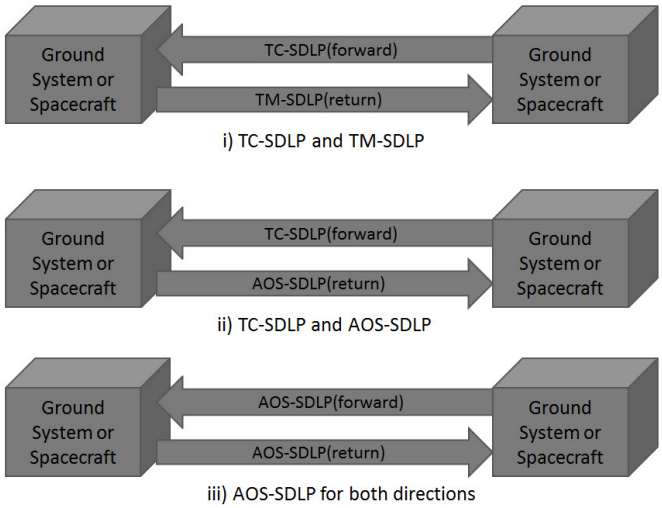


Fig. 5. Different Usable Combination of Space Link Protocols.

Link is used and rovers can respond separately by telemetering data to the Orbiter to be relayed or returned to Earth when a direct-to-Earth link becomes available.

Three combinations of Space Data Link Protocols that are frequently used by space missions are shown in Figure 5.

B. Security Issues and Measures taken

Data Link layer security issues are handled by Link Layer Security which is effective enough to keep this layer safe. The error handling codes used in the supporting layer of Channel Coding are also very helpful to ensure integrity of this layer.

VI. Space Channel Coding layer

Channel Coding is a method by which data can be sent from a source to a destination by processing it in such a way that distinct messages that are created, are easily distinguishable from one another. This allows reconstruction of the data with low error probability, eventually improving the performance of the channel. It is also used to protect the transfer frames against telemetry channel errors induced by noise. The Channel Symbols output by the channel encoder are the basic data units of the CCSDS TM Synchronization and Channel Coding which interface with the physical layer below. *Channel Coding* systems are divided into the four main coding methods:

- Convolutional Code
- Reed-Solomon Code
- Concatenated Code and
- Turbo Codes.

Detail discussion on the different types of Coding is way out of the scope of this paper.

VII. Space Physical Layer

CCSDS has a standard for the Physical Layer called the Radio Frequency and Modulation Systems [9] to be used for space links between spacecraft and ground stations. The purpose of this layer is to provide efficient modulation with specified frequencies to create channels between spacecrafts. The frequency and modulation techniques are different for the different parts of the peer to peer spacecraft communications system. Within the physical layer, there are additional layers covering the technical characteristics, policy constraints, and procedural elements relating to communications services provided by radio frequency and modulation systems. For current networks CCSDS recommended to use **S**, **X**, or **Ka** Band for communication.These bands allocated to Space Research/Deep Space shall only be used by spacecraft engaged in interplanetary research and whose range exceeds a specified distance of 2.0×10^6 km. CCSDS commands to use *Phase Shift Keying* and *Phase Modulation* of the carrier frequency.

The *Reed-Solomon* method is used as the initial error-correcting block code prior to the selected secondary encoding scheme. Again, details of the implementation of these layer is way out of the scope of this paper.

A. Security Issues and Measures taken

Physical Signal security has always been a challenge for security folks. Different modulation techniques as discussed helps to ensure a level of security on the physical layer. More research is on to ensure tighter security on this layer.

VIII. Space End-to-End Security

In the end, we discuss on the security implementation provided by CCSDS Space Security Protocol Layer. Two of such security protocols are Internet Protocol Security (IPSec) and SCPS Security Protocol (SP). The IPSec is used in the Earth side, while other implement deep space end-to-end security protocols [10], [11].

A. System Security Policy (SSP)

While selecting the appropriate security services for a particular mission, the first task is to assess the security threats to the system. This task is normally part of the development of a System Security Policy (SSP). The SSP is usually a formal document providing a description of the system, the top-level security objectives, and identification of the specific information-security and system-availability threats against the system. As part of the SSP development process, a threat assessment should be performed. The threat assessment should assess the vulnerabilities of the system and then establish the likelihood, consequences, and cost of realization of each threat to the system.

B. Threats to Space Systems

The threats to a space mission data system can be categorized into three areas: a) Threats to radio frequency (RF) transmissions b) Threats to the space system resources (e.g., spacecraft and onboard instruments) c) Threats to the ground element of the mission data system. Again, according to behavior of the attacks, they can by of two types:

- Active attacks
- Passive Attacks

In case of an active attack, the attacker has direct access to data or system. On the other hand, a passive attack is typically accomplished by eavesdropping (e.g., listening in on an RF transmission, wiretapping, packet sniffing).

C. Space Data System Reference Model

In the space data system reference model, security measures are distributed among different layers. Figure 6 shows the different protocols used in different layers of the reference model.

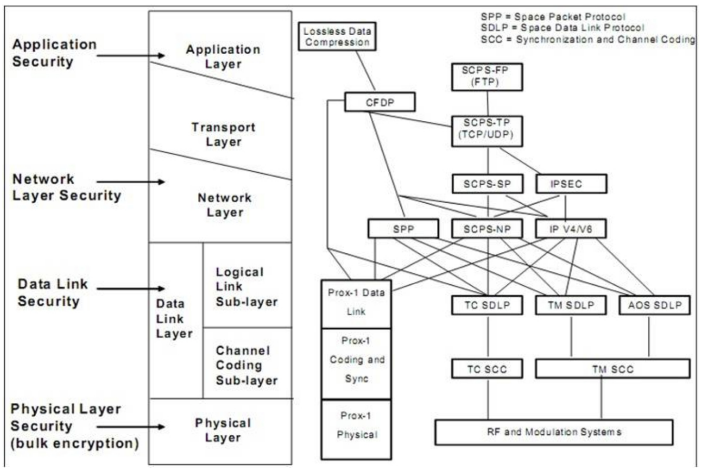


Fig. 6. Security Protocol Distribution in Space Data System Reference Model Layers

D. Security Mechanisms

1) **Confidentiality:** Encryption provides confidentiality service for communications. Any of the following encryption method can be used according to security level requirements:

- Data Encryption Standard (DES) algorithm
- Advanced Encryption Standard (AES)
- Rivest-Shamir-Adleman (RSA) public key system

2) **Authentication:** Data authentication is usually achieved by appending an extra unit of information to the original message. This extra unit of information is called the digital signature, which helps to authenticate the data.

3) **Data Integrity:** To ensure data integrity in Space Communication, the following methods are used:

- Cyclic Redundancy Check (CRC)
- Message Digest 5 (MD5)
- Secure Hash Algorithms (SHA-1, SHA-2)

To ensure data integrity, the receiver generates a corresponding check value by performing an operation (which may be cryptographic) on the data and compares the result to a received value to determine if the data has been modified in transit. In some applications, both authentication and individual data unit integrity can be provided by one mechanism.

IX. Conclusion

This paper is based on the understanding of the current system and protocols used in Interplanetary Network. Different layers of CCSDS (Consultative Committee for Space Data Systems) protocol were covered, which is the recommended by all the Space Communication Organizations including NASA. The resources used for this paper are highly dependable, mainly on informational reports from CCSDS. In addition to that, the insight in security issues of each layer is a unique feature of this paper. Most of the security measures are implementation specific, so protocols and architectures discuss very little about these issues. Hopefully, we shall move to more security centric designs in protocols and architectures in near future.

Acknowledgment

The authors would like to thank the review committee for their efforts and valuable feedback.

References

[1] http://public.ccsds.org/implementations/missions.aspx

[2] http://public.ccsds.org/publications/default.aspx

[3] *CCSDS File Delivery Protocol (CFDP).* Recommendation for Space Data System Standards, CCSDS 727.0-B-4. Blue Book. Issue 4. Washington, D.C.: CCSDS,January 2007.CCSDS 130.0-G-2.

[4] *Space Packet Protocol.* Recommendation for Space Data System Standards, CCSDS 133.0-B-1. Blue Book. Issue 1. Washington, D.C.: CCSDS, September 2003.

[5] *Space Communications Protocol Specification (SCPS)-Network Protocol (SCPS-NP).* Recommendation for Space Data System Standards, CCSDS 713.0-B-1. Blue Book. Issue 1. Washington, D.C.: CCSDS, May 1999.

[6] *TC Space Data Link Protocol.* Recommendation for Space Data System Standards, CCSDS 232.0-B-1. Blue Book. Issue 1. Washington, D.C.: CCSDS, September 2003.

[7] *TM Space Data Link Protocol.* Recommendation for Space Data System Standards, CCSDS 132.0-B-1. Blue Book. Issue 1. Washington, D.C.: CCSDS, September 2003.

[8] *AOS Space Data Link Protocol.* Recommendation for Space Data System Standards, CCSDS 732.0-B-2. Blue Book. Issue 2. Washington, D.C.: CCSDS, July 2006.

[9] *Radio Frequency and Modulation Systems-Part 1: Earth Stations and Spacecraft.* Recommendation for Space Data System Standards, CCSDS 401.0-B-17. Blue Book. Issue 17. Washington, D.C.: CCSDS, July 2006.

[10] *Space Communications Protocol Specification (SCPS)-Security Protocol (SCPS-SP).* Recommendation for Space Data System Standards, CCSDS 713.5-B-1. Blue Book. Issue 1. Washington, D.C.: CCSDS, May 1999.

[11] *Security Threats against Space Missions.* Informational Report,CCSDS 350.1-G-1,Green Book.CCSDS,October 2006.

[12] *Proximity-1 Space Link Protocol-Data Link Layer.* Recommendation for Space Data System Standards, CCSDS 211.0-B-4. Blue Book. Issue 4. Washington, D.C.: CCSDS, July 2006.

Power Pre-emphasised OFDM for RSOA Modulation in High Speed WDM-PONs

P. K. Choudhury
Dept. of Electronics and Communication Engineering
Khulna University of Engineering & Technology (KUET)
Khulna-9203, Bangladesh.
E-mail: pallab@ece.kuet.ac.bd

Abstract— **Orthogonal frequency division multiplexing (OFDM) technique is useful for limited bandwidth reflective semiconductor optical amplifier (RSOA) modulation to design high-speed passive optical networks (PONs). However, the capacity improvement of using conventional OFDM format is severely degraded over the fiber transmission due to the presence of RSOA chirp induced RF power fading. Therefore, we proposed a pre-emphasis OFDM for RSOA modulation to enhance the system performance against both RSOA low bandwidth and high chirp parameter. Experimental results showed that the transmission speed is reached up to 17.1 Gb/s at the bit error rate (BER) within forward error correction (FEC) limit over 25 Km fiber span by using commercially available RSOA of bandwidth < 1 GHz.**

Keywords— Optical fiber communications, Orthogonal frequency division multiplexing (OFDM), Reflective semiconductor optical amplifiers (RSOAs), Passive optical networks (PONs).

I. INTRODUCTION

To meet the growing bandwidth demand from end user's broadband services, Wavelength Division Multiplexing Passive Optical Network (WDM-PON) is a promising access solution. For the cost-effective development of such network, it is desirable to design a 'colorless' Optical Network Units (ONUs). Among the different proposed access architectures, Reflective Semiconductor Optical Amplifier (RSOA) based WDM-PONs is the most acceptable one [1][2]. The key device of this system is the RSOA itself, due to its wide optical bandwidth, re-modulation, optical amplification and integration capabilities. However, RSOA has low electro-optical modulation bandwidth. To design high speed RSOA based WDM-PONs, the Orthogonal Frequency Division Multiplexing (OFDM) modulation has recently been used in several investigations [3][4]. The OFDM modulation shows number of advantages, for example, its inherent robustness against chromatic dispersion, the flexibility to allocate data among different subcarriers and above all, high spectral efficiency that promotes this modulation technique to use in limited bandwidth devices like RSOA.

However, OFDM modulated optical signal is affected by power fading in fiber transmission, which is again enhanced due to the non-negligible chirp factor of RSOA. Thus, it is necessary to consider both limited bandwidth and chirping effect of RSOA to design optical OFDM based high speed WDM-PONs. In [4], transmission capacity was improved by OFDM modulation based on bit-power loading (BPL) algorithm, where the investigation was focused on limited bandwidth operation of RSOA. To overcome the implementation complexity of BPL, more simplified approach called as pre-emphasised OFDM modulation was utilized for conventional direct detection optical system [5]. However, none of the above investigations was considered the modulator chirp, which has significant impact on transmission performance for optical OFDM system [6].

Therefore, this work highlights number of design issues for chirped RSOA modulation by using OFDM signal, which include; i) A theoretical study to determine the chirp parameter and effective modulation bandwidth for the RSOA under use. This information is necessary to select the useful OFDM signal baseband bandwidth by avoiding the fading null frequency ii) Identify the maximum transmission capacity of the system by using the proposed power pre-emphasis technique. Moreover, determine the optimize operating conditions of RSOAs, for example, the seeding power of RSOA has a significant impact to reach high data rate [7] and also it is related with the power budget of the system. Therefore, an optimized value need to be identified to satisfy the above mentioned criteria. iii) Finally, explore the dependency of system maximum capacity with three important parameters, i.e. seeding power of RSOA, receiver sensitivity and FEC limit. This information could be interesting for system designer to check capacity limit with various operating conditions. Highlighting the above mentioned aspects and make a theoretical with experimental analysis are the main focus of this paper.

The paper is organized as follows. Section II provides a theoretical study of fading effect from RSOA chirp. Section III presents the experimental setup. Section IV discusses the system performance by using pre-emphasis techniques. Finally, Section V summarizes the paper.

II. RSOA CHIRPED INDUCED FADING EFFECT

In RSOA, due to the phase-gain coupling, the modulated signal is affected by frequency chirp, which is governed by the following expression [8]:

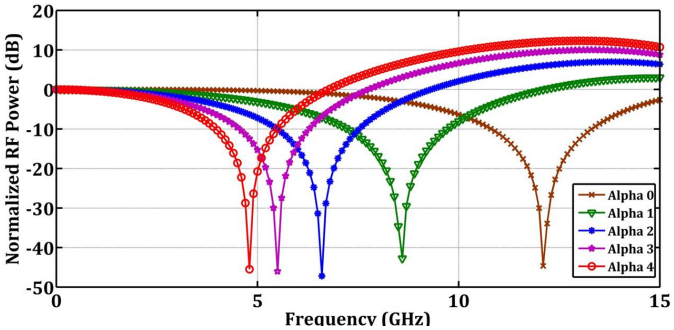


Fig.1. Normalized received RF power for different values of alpha parameters

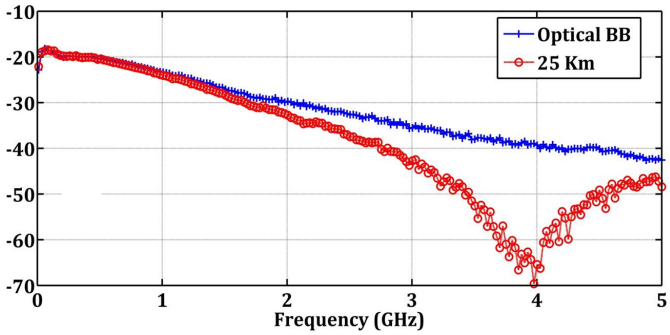


Fig.2. Experimentally measured frequency response of RSOA for optical back-to-back and 25 Km fiber transmission

$$\phi(t) = \frac{-\alpha_{eff}}{2} \ln(\Delta G(t)) \qquad (1)$$

Where $\phi(t)$ is the phase modulation, α_{eff} denotes the effective linewidth enhancement factor (i.e. the chirp parameter), ΔG is the gain variation induced by direct current modulation. Therefore, overall phase variation depends on both chirp parameter and the gain transition of the device. For the multicarrier modulation format like OFDM, this phase variation is enhanced due to the increase of gain transition from the high value of peak-to-average power ratio (PAPR).

On the other hand, direct modulation of an optical carrier results a double-sideband signal where the baseband frequencies are located in either side of optical carrier. When this modulated signal propagates along with the dispersive fiber, the two optical sidebands experience a different phase shift with respect to carrier. After photo detection, the subcarriers RF power will be varied depending on the relative phase variation among them. Without considering the fiber loss and chirp parameter of the modulator, the RF power fading can be represented as [9]:

$$P_{rf} \propto \cos^2(2\pi^2 \beta_2 L f^2) \qquad (2)$$

Where β_2 is the group velocity dispersion (GVD) parameter, L is the fiber length, and f represents the frequency subcarrier of electrical OFDM signal. However, RF fading power in (2) can be modified by considering the modulator chirp parameter [10]:

$$P_{rf} \propto (1+\alpha^2)\left|\cos^2(2\pi^2 \beta_2 L f^2 + \tan^{-1}\alpha)\right| \qquad (3)$$

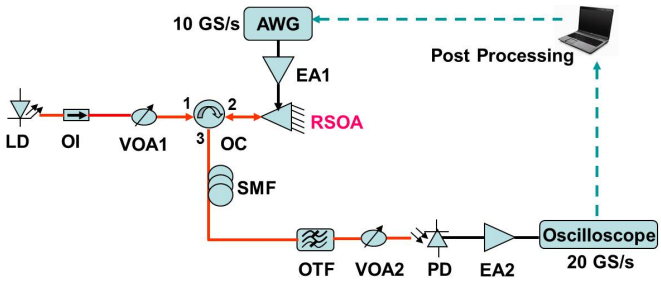


Fig. 3. Experimental setup. LD: Laser Diode, OI: Optical Isolator, VOA: Variable Optical Attenuator, OC: Optical Circulator, EA: Electrical Amplifier, OTF: Optical Tunable Filter, PD: Photo Diode

Where, α is the chirp parameter. To quantify the severity of chirp parameter on RF power fading, Fig.1 shows the normalized (with respect to DC) RF power as a function of modulation frequency at different values of alpha parameter for the transmission over 25 Km fiber span. The chirp parameter has significant influence on the signal fading condition. In particular, the null frequency is moved towards the DC component while the alpha parameter is increased. And thus, the system 3 dB bandwidth is reduced due to presence of large value of device chirp.

Beside the above theoretical analysis, an experimental investigation was also performed by using the RSOA to measure its chirp induced power fading. Figure 2 shows the experimentally measured RF power for a directly modulated RSOA in both optical back to back and fiber transmission over the distance of 25 Km. Due to the high value of RSOA chirp, the system shows the fading dip near 4 GHz even in such short fiber transmission. Moreover, with the value of RSOA first null frequency at 4 GHz, chirp parameter can be measured to 5.67 by using (3), which is a reasonable value with respect to previous studies [11]. Beside this RSOA chirping effect, the device also shows the limited modulation bandwidth < 1GHz as illustrated in fig. 2.

Therefore, high speed (>10 Gbps) access system based on OFDM signal for RSOA modulation is fundamentally limited by chirp parameter and modulation bandwidth of the reflective device itself. To overcome the aforementioned limitations, we proposed and also experimentally investigated an adaptive OFDM format based on pre-emphasis power with selective bit loading for each subcarrier, which will be addressed in details in section IV.

III. Experimental Setup

In order to evaluate the performance for RSOA based PONs system by using OFDM, the experimental setup was realized as illustrated in Fig. 3. A CW optical source was used at 1550 nm with an output optical power of 7 dBm. Followed by optical source, there were also an optical isolator (OI), a variable optical attenuator (VOA) and an optical circulator (OC). The VOA was used to control the input seeding power of RSOA, whereas the OC was used to separate the incident and modulated optical signals to and from RSOA. The modulated signal was passed through the 25 Km single mode

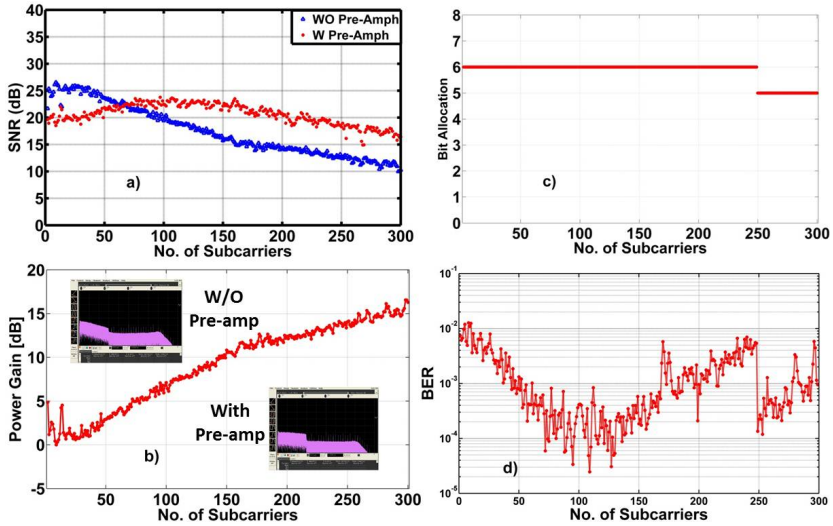


Fig. 4. a)Measured channel SNR for without and with pre-emphasis power b) Applied pre-emphasis power c) Bit allocation corresponding to maximum capacity (Insets show the OFDM spectrum with or without pre-emphasis) d) BER for each subcarrier.

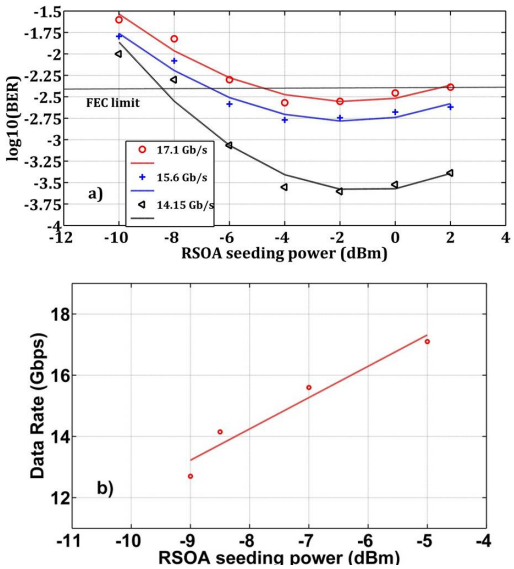


Fig.5. a) BER performance as a function of RSOA seeding power for different data rates; PD input power= -5.5 dBm b) Data rate against RSOA seeding power for the BER=3.8e^{-3}.

fiber (SMF). The bias current of RSOA was set as 60 mA with a temperature of 20º C. Under this condition, the measured 3 dB bandwidth was 0.96 GHz (see fig.2). The device has small signal gain of 20 dB and the saturated output power of 3 dBm. The OFDM signal was constructed in offline and then uploaded to a 5 GHz, 10 GS/s Arbitrary Waveform Generator. The signal was composed of 512 subcarriers (the first carrier has no power). The real value OFDM signal was generated by taking the complex conjugate of 512 subcarriers. The subcarrier spacing and total symbol time were 9.77 MHz and 106 ns respectively.

To control the power and modulation order of each carrier, two approaches were adapted; firstly, based on a bit-power loading algorithm and secondly, use of pre-emphasis power with selective bit loading technique. The details of these techniques will be discussed in the corresponding sections. Finally, the RSOA was directly modulated by OFDM signal by using an electrical amplifier.

On the other side, the optical receiver was composed by an optical tunable filter (OTF) of 25 GHz bandwidth, a VOA and a 14 GHz PIN photodiode. Once received, the electrical signal passed through an electrical amplifier and sampled by a 20 GS/s real time digital oscilloscope. Due to record memory length limitations, only four hundred OFDM symbols were captured at this sampling rate. The acquired samples were then processed offline for synchronization, equalization and bit error rate measurement.

IV. Results and Discussion

The chirped Optical OFDM signal with pre-emphasis coefficients can be expressed by [3]:

$$I \cong 1 + 2\sqrt{1+\alpha^2}\sum_{n=1}^{N}\kappa_n x_n \cos(n\omega t + \theta_n)\cos(n^2\theta_d - \theta_\alpha)$$

Where, α is the chirp parameter, x_n and θ_n are the amplitude and phase of nth subcarriers, θ_d and θ_α are the phase variation due to fiber dispersion and modulator chirp respectively. κ_n is the pre-emphasis power coefficients for n^{th} subcarriers, can be evaluated as inverse of measured channel SNR after fiber transmission i.e. $\kappa_n = \mathrm{SNR}_n^{-1}$. Figure 4(a) shows the measured SNR for 25 Km fiber transmission with and without the use of power pre-emphasis. By applying the pre-emphasis power, the subcarriers SNR is significantly improved by 5 dB especially in the region of high frequency components to allow higher-level modulation format for capacity improvement. Note that this SNR performance results from both the limited bandwidth response of RSOA and fading effect. And thus, the pre-emphasis OFDM signal is useful to minimize both the effects with improved SNR.

Pre-emphasis technique requires optimize allocation of subcarrier power due to the limitation imposed by resolution bits/dynamic range from the commercially available digital to analog sampler. Here we selected the OFDM electrical signal bandwidth of 2.93 GHz (~ 300 subcarriers), which gives around 15 dB power variation from 1st to N^{th} subcarrier, as shown in Fig. 4(b), a tradeoff between the generation of sampler quantization noise and the addition of pre-emphasised power to equalize the channel response. Figure 4(c) illustrates the bit allocation in each carrier, showing that the modulation order was varied from 64 QAM to 32 QAM depending on channel SNR with the limit of required BER=3.8e^{-3}[12][13]. The BER performances for different subcarriers were evaluated from the computation of Error Vector Magnitude [14]. The aggregate BER was < 2e-3 which is below the FEC limit as shown in Fig. 4(d). Under this condition, the system gives the maximum capacity of 17.1 Gb/s even with 25 Km fiber transmission, showing a significant improvement by applying the proposed OFDM modulation. In addition, the insets of Fig. 4(b) show the screen images of OFDM spectrum that confirm the effect of channel equalization with/without the application of pre-emphasis technique.

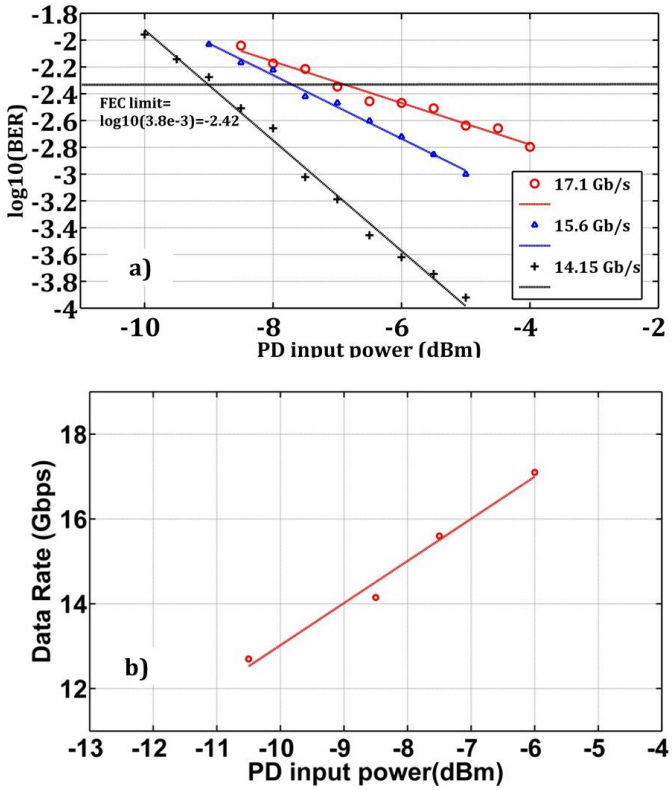


Fig. 6: a) BER performance in terms of receiver sensitivity for different data rates; RSOA seeding power = -2 dBm, b) Data rate against receiver sensitivity for thr BER=3.8e^{-3}

For RSOA based PONs, seeding power is a crucial parameter as it has the impact on system power budget, transmission capacity and also on E/O modulation characteristics of the device itself. Hence, Fig. 5(a) shows an experimental characterization of RSOA seeding power with respect to BER performance by considering different target data rate i.e. 17.1 Gbps, 15.6 Gbps, 14.15 Gbps. There was always an optimum seeding power to achieve a low BER. However, in either side of the optimum value, BER was increased due to noise effect in low seeding power and also for low frequency cutoff when the device is highly saturated with high seeding power. Moreover, Fig. 5(b) illustrates the achievable data rate for various levels of RSOA seeding power to maintain the target BER. System capacity can be improved by sending high label of RSOA seeding power. However, due to the limitation imposed by system power budget, RSOA input power can not be increased arbitrarily. In particular, for 17.1 Gb/s data transmission, downstream power budget was around 12 dBm whereas, for other two data rates i.e. 15.6 Gbps and 14.15 Gbps, the power budget was increased to 14 dBm and 16 dBm respectively.

Based on the optimum RSOA seeding power of -2 dBm, the system performance can also be analyzed by photo diode input power as a function of BER for different data rates, as shown in Fig.6(a). To achieve the maximum capacity, the required PD input power was -6 dBm to maintain the BER within FEC limit. The receiver sensitivity can be improved another 3 dB if the target data rate is 14.15 Gbps. The dependency between

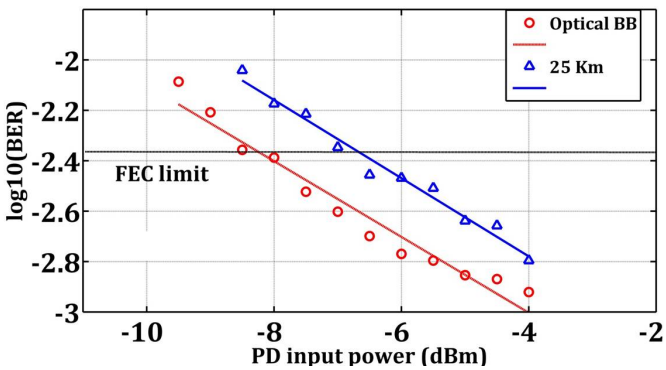


Fig. 7: BER performance of optical back to back and 25 Km fiber transmission for the data rate 17.1 Gb/s.

data rate and photodiode received power were also shown in Fig. 6(b). This performance study indicates the necessity of sufficient input received power in order to maintain the BER <1e^{-4} and also to maximize the system capacity, however, this condition could have worst impact on upstream power budget. Alternatively, a receiver with APD or pre-amplification might be a good choice to improve this dilemma condition [15] where the system more focus on capacity maximization rather than the cost minimizations.

Finally, Fig. 7 shows the BER performance for optical back to back and over 25 km fiber transmission for the case of maximum system capacity and by maintaining -2 dBm RSOA seeding power. There was around 1.5 dB penalty between them. The possible reason for this penalty due to the subcarrier-subcarrier intermixing product which has strong impact in the presence of modulator chirp and fiber dispersion [3]. To overcome this penalty, a guard interval (GI) [16] could be placed between optical carrier and OFDM signal. However, it will occupy a useful portion of the available bandwidth of the modulator and moreover, by using RSOA, existence of GI in OFDM generation may not be a good choice as the system is already limited by reduced frequency response of the device itself.

V. Conclusion

We proposed and experimentally characterized an adaptive OFDM modulation technique by using pre-emphasis power and selective bit loading for RSOA based next generation high speed (>10Gb/s) WDM PONs. It showed that the pre-emphasis power loading technique provides an optimal solution in the presence of both RSOA limited bandwidth and chirped induced fading effect. Transmission experiment was carried out over 25 Km fiber span to achieve the system capacity >10Gb/s. Furthermore, the work was also highlighted the dependency of system capacity on various operating conditions i.e. seed power, received power and BER limit, in a RSOA based high speed next generation WDM-PONs.

ACKNOWLEDGMENT

The author would like to acknowledge the support from Scuola Superiore Sant'Anna (ITALY) for conducting the experimental part of this work and HEQEP project from University Grant Commission of Bangladesh under the grant contract CP-2378.

REFERENCES

[1] E. Wong, "Next-generation broadband access networks and technologies," *J. Lightwave. Technol.*, vol. 30, no. 4, pp. 597–608, Feb. 15, 2012.

[2] J. Prat, C. Arellano, V. Polo, and C. Bock, "Optical network unit based on a bidirectional reflective semiconductor optical amplifier for fiber-to-the-home networks," *IEEE Photon. Technol. Lett.*, vol. 17, no.1, pp. 250–252, Jan. 2005.

[3] T. Duong, N. Genay, P. Chanclou, B. Charbonnier, A. Pizzinat, R. Brenot, " Experimental demonstration of 10 Gbit/s upstream transmission by remote modulation of 1 GHz RSOA using Adaptively Modulated Optical OFDM for WDM-PON single fiber architecture," in *Proc. ECOC 2008*, Paper Th3.F1.

[4] M-K. Hong, N. C. Tran, Y. Shi, E. Tangdiongga, S-K. Han, A. M. J. Koonen, "10Gb/s Transmission over 20km Single Fiber Link using 1GHz RSOA by Discrete Multitone with Multiple Access," in *Proc. ECOC 2011*, Paper Th11.C3.

[5] Y. Gao, J. Yu, J. Xiao, Z. Cao, F. Li, and L. Chen, "Direct-Detection Optical OFDM Transmission System With Pre-Emphasis Technique," *J. Lightw. Technol.*, vol. 29, no. 14, pp. 2138-2145, Jul. 2011.

[6] Chia-Chien Wei, "Small-signal analysis of OOFDM signal transmission with directly modulated laser and direct detection," *Optics Letters*, vol. 36, no. 2, pp. 151-153, Jan. 2011.

[7] C.W. Chow, C.H. Yeh, Y.F. Wu, H.Y. Chen, Y.H. Lin, J.Y. Sung, Y. Liu and C.-L. Pan, "13 Gbit/s WDM-OFDM PON using RSOA-based colourless ONU with seeding light source in local exchange," *Electronics Letters*, vol. 47, no. 22, pp. 1235-1236, Oct. 2011.

[8] L. Occhi, L. Schares, and G. Guekos, "Phase Modeling Based on the α-Factor in Bulk Semiconductor Optical Amplifiers," *IEEE J. of Selected Topics in Quantum Electronics*, vol. 9, no. 3, pp. 788-797, May/June 2003.

[9] H. Schmuck, "Comparison of optical millimeter-wave system concepts with regard to chromatic dispersion," *Electronics Letters*, vol. 31, no. 21, pp. 1848–1849, Oct 1995.

[10] F. Devaux, Y. Sorel, and J.F. Kerdiles, "Simple Measurement of Fiber Dispersion and of Chirp Parameter of Intensity Modulated Light Emitter," *J. Lightw. Technol.*, vol. 11, no. 12, pp. 1937-1940, Dec. 1993.

[11] I. Cano, M. Omella, J. Prat, P. Poggiolini, "Colorless 10Gb/s extended reach WDM PON with low BW RSOA using MLSE," ," in *Proc. OFC 2010*, Paper OWG2.

[12] N. Cvijetic, M. F. Huang, E. Ip, Y. K. Huang, D. Qian, T. Wang, "1.2 Tb/s Symmetric WDM-OFDMA-PON over 90km Straight SSMF and 1:32 Passive Split with Digitally-Selective ONUs and Coherent Receiver OLT," in *Proc. OFC* 2011, Paper PDPD7.

[13] ITU-T Recommendation G.975.1, Appendix I.9, 2004.

[14] D. Hillerkuss, R. Schmogrow, T. Schellinger, M. Jordan, M. Winter, G. Huber, T. Vallaitis, R. Bonk, P. Kleinow, F. Frey, M. Roeger, S. Koenig, A. Ludwig, A. Marculescu, J. Li, M. Hoh, M. Dreschmann, J. Meyer, S. Ben Ezra, N. Narkiss, B. Nebendahl, F. Parmigiani, P. Petropoulos, B. Resan, A. Oehler, K. Weingarten, T. Ellermeyer, J. Lutz, M. Moeller, M. Huebner, J. Becker, C. Koos, W. Freude, and J. Leuthold, "26 Tbit/s line-rate super-channel transmission utilizing all-optical fast Fourier transform processing," *Nature Photonics* 5, pp. 364–371, June 2011.

[15] A. Gharba, P. Chanclou, J. Le Masson, M. Ouzzif, N. Genay, B. Charbonnier, E. Grard, V. Rodrigues and M. He´lard, "Optical transmission performance in AMOOFDM downlink NG-PON systems using APD and pin receivers," *Electronics Letters*, vol. 47, no. 13, pp. 761-762 , June 2011.

[16] A. J. Lowery and J. Armstrong, "Orthogonal-frequency-division multiplexing for dispersion compensation of long-haul optical systems," *Opt. Express*, vol. 14, no. 6, pp. 2079–2084, Mar. 2006.

A New Background Updating Model for Motion Detection Considering Future Frame

Arifur Rahaman, Md. Mehedi Hasan,Riaz Ahmed, Mirza Md. Shahriar Maswood, Md. Mostafizur Rahman
Dept. of Electronics & Communication Engineering,
Khulna University of Engineering & Technology,
Khulna-9203, Bangladesh
{arifur_ece, mehedi_ece}@yahoo.com, {riazahmed0111, shaonkuetece05}@gmail.com, mostafiz963@yahoo.com

***Abstract*— Motion detection is now very much popular for its different type of applications. And the most important thing for motion detection is the background updating procedure. For background updating, all the existing work is done actually depend on the present and past frame. In this paper we proposed a new background updating model for motion detection which mainly differs by considering the future frame. This work is done by taking a number of frames initially to find median from these frames and stored as the past background. Then some frames are taken as future frames and computes a future background from these frames by finding the median. Now if motion occurs in any frame which lies between the past background and this updated future background can be detected easily and accurately by comparing the frames with both of these two updated background. And after some moment the updated future background will be the past background and again some of the frames will be considered as the future frames to calculate the new future background. Our proposed method can update the background accurately as well as can detect motion or moving object perfectly. We have applied this method on various real time data and got promising results.**

Keywords-background updating; past and future background; motion detection; object detection.

I. INTRODUCTION

Motion detection is a process through which the moving object can be detected. This is actually done by finding the pixel value difference from frame to frame. This motion detection algorithm is used in many video surveillance systems for positive alarm of motion. Many restricted area where unauthenticated object should not be present there it has a great use. For the detection or observation of criminal activity video surveillance system is now used almost every crowded place. This has a wide use in military section. The electronic motion detection is basically based on the difference between pixel values in two consecutive video frames and that's why it is able to locate the motion when there is any change in pixel values. This is the simplest way for motion detection but for detecting motion more accurately there has been done many improved work in previous days.

For motion detection there are mainly three basic methods present, first is background subtraction, the other is temporal differencing and the last is optical flow [1]. In temporal differencing the video frame is separated in a time scale and then find if there is any change [2]. Another method for motion detection is optical flow. It is a vector based method which make point based analysis on multiple frames for motion detection [3]. Background subtraction is very popular for static background applications. Much work has been done in the basis of background subtraction [4], [5], [6]. One work showed the observation of motion in static camera scene, which is the main task in visual surveillance [7]. Also another work has done for automatically discovering key motion patterns happening in a scene by observing the scene for an extended period [8]. For 3D motion tracking motion detection is also important and for this purpose model based motion tracking system where distributed networks camera is used has been developed [9]. Also motion is detected along with video quality analysis for video surveillance systems in where three modules is used- a background modeling (BM) module, an alarm trigger (AT) module, and an object extraction (OE) module [10]. Motion detection is done for crowed analysis [11], on road vehicle detection systems [12] and intelligent distributed surveillance systems [13] etc. Motion detection is also applied for observation of object behavior [14] and observation of human behavior for homeland security applications [15]. One of our previous works has done on the basis of temporary background selection and then develops a permanent background through which motion can detect perfectly except some exceptional cases [16].

In this work we proposed a new method in where we consider number of future frames to model future background. The major feature is the past background and future background. Initially from some video frame the median is calculated and then these median pixel values are stored as the past background. After storing the past background then consider some frames as future frames which are lies ahead of time than the past. From these future frames another background is stored by finding the median of these frames. Now there remain some frames between the past background and the future background. If any positive motion occurs on these frames then simply by comparing with both the past background and future background this can be detected. Only if there finds any pixel value difference there a motion is detected. And for updating after a certain time instance it can update automatically. After some time the future background will be the updated past background and the updated future background will be calculated as explained before. In this method chance of a false positive alarm is very low. We have applied this method in different real time data and we observed promising results from this method.

The rest of the paper organized as follows. Section II briefly overviews of our proposed framework. Section III describes the proposed algorithm of our method. Experiments and result analysis are described in Section IV. Section V shows the discussion of this work. We conclude the paper in Section VI.

II. OVERVIEW OF OUR FRAME WORK

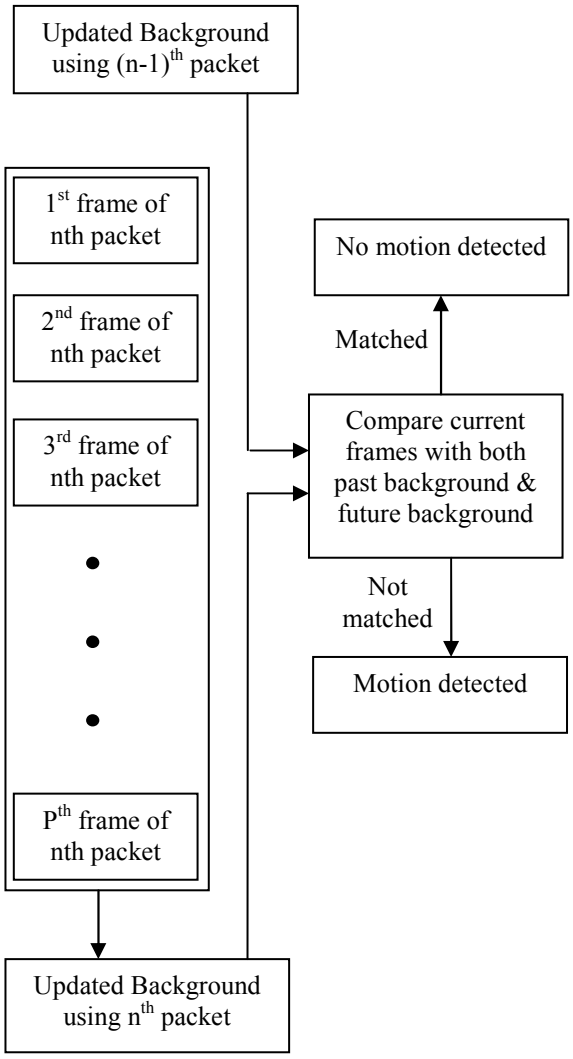


Fig 1. Block diagram for background modeling and motion detection

Fig. 1 describes the block diagram of the background modeling process of our proposed method. In this method we have initially updated a background from first 40 frames which lies in the $(n-1)^{th}$ frame. And then with the next 40 frames from n^{th} packet another background is updated that we have defined as future background. Then with these two updated background the frames which lies on the nth packet is compared for motion detection. If any pixel values of these frames are mismatched with both of these two backgrounds then there a positive motion will be detected. If any frame's pixel value matches with any of the background then there will be no motion detected. After this when $(n+1)^{th}$ packet comes then the nth will be the updated past background and the new future background will be updated from the $(n+1)^{th}$ packet. Here one of the backgrounds defined as the future background because the frames which are compared for motion detection comes first than the updated background in time scale. Mainly the future background comes later than the frames being compared. That's why here we have defined it as future background.

III. ALGORITHM OF OUR PROPOSED METHOD

1. Select a certain number of frames for making each packet.
2. For updating background and motion detection:

   Step 1) Make an initial or past background by using first packet ($(n-1)^{th}$ packet) which contains initial frames.

   Step 2) Create a future background by using the next packet (n^{th} packet).

   Step 3) Compare the nth packet's frames with both past background and future background consequently.

   If both □ pixel value of any frame- pixel value of past background □ > threshold and □ pixel value of any frame- pixel value of future background □ > threshold then there detects a motion.

   Otherwise there is no motion occurs.

   Step 4) Now if the $(n+1)^{th}$ packet's all frames acquired then update the current past background with the previous future background and update the current future background by using $(n+1)^{th}$ packet's frames.

   Step 5) Go to **Step 3** and set n^{th} packet= $(n+1)^{th}$ packet.

3. If no more frames available for updating future background then **Stop**.

IV. EXPERIMENTAL STUDY

A. Acquisition of Image

The acquisition of video image was performed with the camera sth-dcsg-9cm. The STH-DCSG is a revolutionary stereo head with a global shutter specially designed for machine vision tasks of scenes with motion, Pre-calibrated so it is plug-and-play, Global shutter, low-noise, high-sensitivity CMOS imager, Color or monochrome, uncompressed video at VGA resolution (15 fps), Includes miniature lenses and SRI Small Vision System software for real-time stereo analysis, Offering high performance and low power with motion capture in a compact package. Here the camera is interfaced with a video surveillance system. When the camera is calibrated, it will produce dynamic video images.

Fig 2. Pictorial view of a STH-DCSG – 9CM camera

B. Experimental Results

1. Indoor Small Object

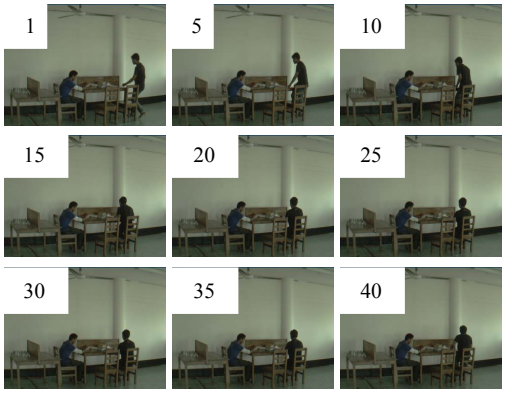


Fig 3. 1st packet of frames for indoor small object (frame1-40)

Fig 4. Past background after first packet of frames acquired

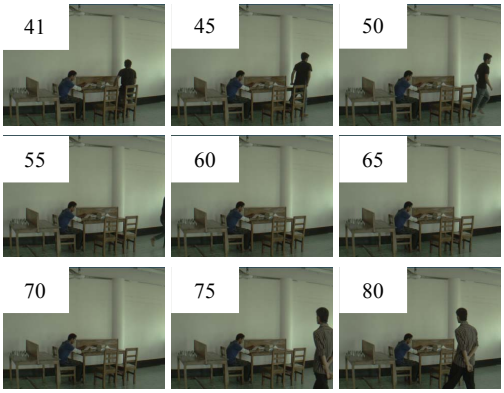


Fig 5. 2nd packet of frames for indoor small object (frame41-80)

Fig 6. Detected future background

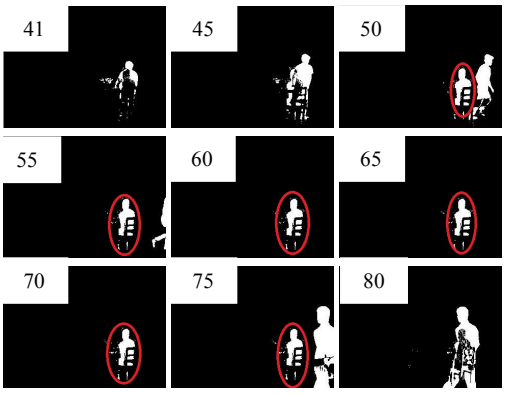


Fig 7. Motion detection using conventional process

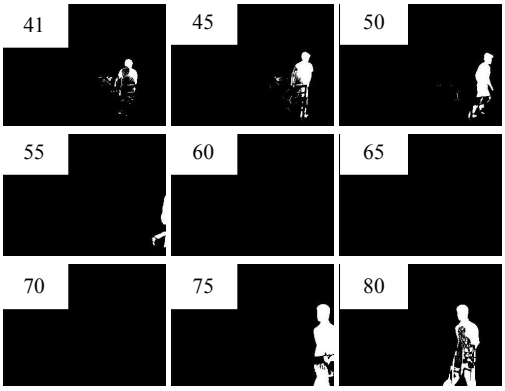


Fig 8. Motion detection using our proposed method

At first we consider indoor small object in real time data to observe the performance of our proposed method. First we consider the scenario which described on fig. 3. Here one object is reading in the reading room and another object is coming towards the first object and sits beside him, which occurs from frame 1 to frame 40. But in frame 41 the second object is started to leave the place and in frame 55 the second object is almost gone. From frame 56 to 70 the first object is only object present there without any motion. In frame 75 third object's entrance is occurred. And the only moving object is there is this third object which moves frame 75 to frame 80. Now in conventional process from the first 40 frames a background is updated for motion detection using the median from the frames. Then when this background is compared with the frames which come next to it, this method detects the stationary object as moving object. When the 1st background compared with frame 41, 45, 50 etc then there it finds the second object as moving because the pixel values differ from these frames. In the frame 65, there is no moving object but the conventional method still shows the 2nd object as a moving object which has no existence on this frame. The erroneous motion detection is shown on the fig. 7 by using a red circle. For this type of difficulty our proposed method reduces this false positive alarm. In our method first 40 frames (frame 1-frame 40) are taken to update the initial background which is shown on the fig. 4 and then the next 40 frames (frame 41-frame 80) are considered as future frame and then from these frame the future background is updated which is also shown in fig. 6. For this reason if we want to find the motion from frame

41 to frame 80 then these frames are compared both with the initial background and with the future background, if pixel value mismatches for both background then it detects a moving object there. The moving object detection by our proposed method is shown in fig. 8, where only the moving object is detected perfectly. And for next upcoming 40 frames the current future background will be the initial or past background and the future background will be updated from next 40 frames. In this process the motion detection or moving object detection is done perfectly.

2. *Outdoor Small Object*

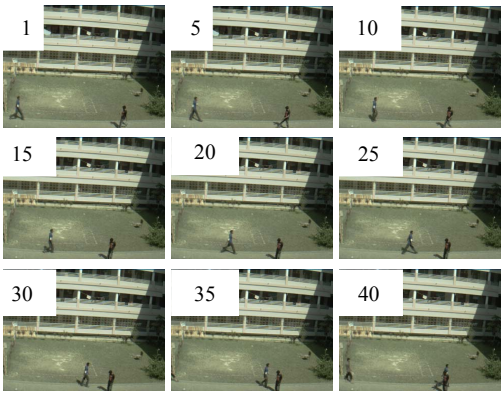


Fig 9. 1st packet of frames for outdoor small object (frame1-40)

Fig 10. Background after first packet of Frames

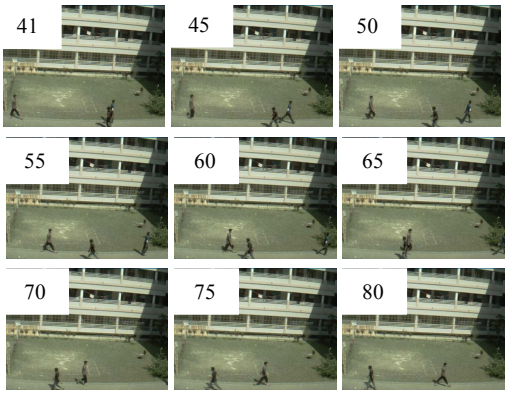


Fig 11. 2nd packet of frames for indoor small object (frame41-80)

Fig 12. Detected background

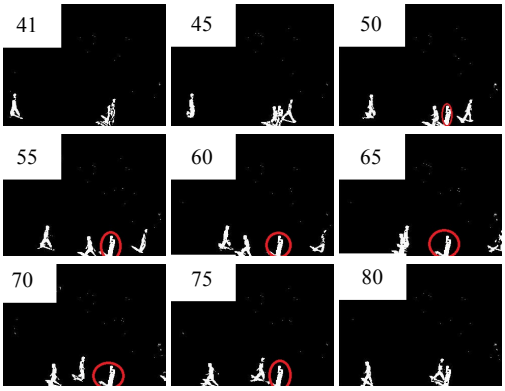


Fig 13. Using normal process

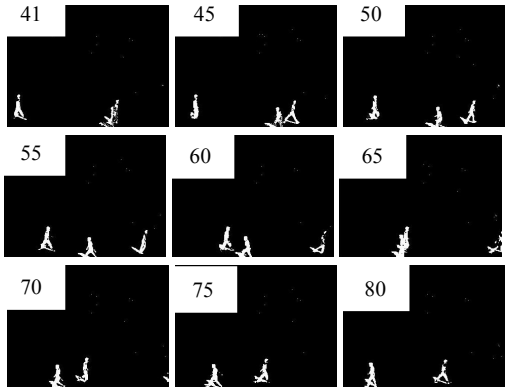


Fig 14. our proposed method

Here we have considered the scenario for outdoor small object. In this case two objects are walking in the road. The object which comes from the right side walks a little and then in a point it stands still and does not have any movement. But the object which enteres from the left side it is continuously walking and crossing the first object. When crossing the man, another object also starts to enter in the road. This whole situation is described in fig. 9 through frame 1 to frame 40. At frame 45 the standing man started to move and frame 65 the first and third object crossed each other. Through frame 66 to 80 they walked away. This is described in the fig. 11. Now in the conventional background updating method from the first 40 frame a background will be updated by calculating the median from the frames. Then for detection of motion this background will be compared with the next coming frames, if it finds any difference on the pixel values then there it detects a motion. The initially updated background is shown on fig. 10. Let's see what's if we only consider this background to detect motion. When this background compared with frames 41 to 80 then in frame 50 it detects an extra object in the frame as moving object which is shown by a red circle. And this false positive alarm is appeared through the next frames. This erroneous picture is shown on the fig. 13. Now if we consider our proposed method in where a future background is found from next 40 frames (frame 41-frame 80) along with the past or initial background. And frames lies between these two

backgrounds are considered for motion detection. Each frame is compared with these two backgrounds and if pixel value mismatches for both background then it detects a moving object there otherwise there will be no moving object. After taken another 40 frames then the current future frame will be the updated past background and the future background will be updated by the newly taken frames. The experimental results by our proposed method is shown in fig. 14 where only the moving object is detected perfectly and this method shows no false positive alarm.

V. DISCUSSION

In this work our target is to develop a method through which motion can be detected perfectly. For this purpose we have considered two backgrounds one is past and another is future. And the frames which lies between the past background and future background is compared with both of these backgrounds to ensure the motion detection accuracy. We have applied this method on two real time video data and observed the performance. Our method has perfectly detected the motion when present in any frame but the traditional method has shown false positive alarm in some cases. Our proposed method may a little bit slow process because of updating the future background but in case of accuracy it shows high perfection.

VI. CONCLUSION

Motion detection is now very much popular due to its dynamic application. Many video surveillance systems use this feature for detection of motion and identification of object. It is also applied in traffic and human behavior recognition. So for all types of applications it is firstly recommended to detect the motion perfectly. And for motion detection by electronic way the background updating process is the most popular and important. Our proposed algorithm is tested on two real time data. Experimental study shows both the traditional method's motion detection result and our method's motion detection result. In some cases traditional method shows a moving object although there is no moving object present there. But our method removes this limitation and only shows the moving object if there is any. So in term of performance our proposed method shows better accuracy than the traditional method. It is also notable that our experimental results show better performance in noisy and open air environment.

ACKNOWLEDGMENT

Authors of this paper wish to gratefully acknowledge the project under HEQEP which is CP-2378 for the financial assistance of this work.

REFERENCES

[1] W. Hu, T. Tan, L. Wang, and S. Maybank, "A survey on visual surveillance of object motion and behaviors," IEEE Trans. Syst., Man, Cybern. C, Appl. Rev., vol. 34, no. 3, pp. 334–352, Aug. 2004.

[2] F.-Y. Hu, Y.-N. Zhang, and L. Yao, "An effective detection algorithm for moving object with complex background," in Proc. IEEE Int. Conf. Mach. Learn. Cybern., 2005, vol. 8, pp. 5011–5015.

[3] M. Black and P. Anandan, "The robust estimation of multiple motions: Parametric and piecewise smooth flow fields," Comput. Vis. Image Underst., vol. 63, no. 1, pp. 75–104, Jan. 1996.

[4] A. Manzanera and J. C. Richefeu, "A robust and computationally efficient motion detection algorithm based on Σ-Δ background estimation," in Proc. ICVGIP, 2004, pp. 46–51.

[5] L. Maddalena and A. Petrosino, "A self-organizing approach to background subtraction for visual surveillance applications," IEEE Trans. Image Process., vol. 17, no. 7, pp. 1168–1177, Jul. 2008.

[6] M. Oral and U. Deniz, "Center of mass model: A novel approach to background modeling for segmentation of moving objects," Image Vis. Comput., vol. 25, pp. 1365–1376, Aug. 2007.

[7] Imran Saleemi, Lance Hartung and Mubarak Shah, "Scene understanding by statistical modeling of motion patterns," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2069-2076, Jun. 2010.

[8] Yang Yang, Liu Jingen and M. Shah, "Video scene understanding using multi-scale analysis," 2009 IEEE 12th International Conference on Computer Vision (ICCV), pp. 1669-1676, 2009.

[9] Yumi Iwashita, Ryo Kurazume, Takamitsu Mori, Masaki Saito, and Tsutomu Hasegawa, "Model-based motion tracking system using distributed network cameras," 2010 IEEE International Conference on Robotics and Automation, pp. 3020-3025, May 2010.

[10] Shih-Chia Huang, "An advanced motion detection algorithm with video quality analysis for video surveillance systems," IEEE Transactions on Circuits and Systems for Video Technology, vol. 21, pp. 1-14, Jan. 2011.

[11] B. Zhan, N. D. Monekosso, P. Remagnino, S. A. Velastin, and L.-Q. Xu, "Crowd analysis: A survey," Mach. Vis. Appl., vol. 19, no. 5/6, pp. 345–357, Sep. 2008.

[12] Z. Sun, G. Bebis, and R. Miller, "On-road vehicle detection: A review," IEEE Trans. Pattern Anal. Mach. Intell., vol. 28, no. 5, pp. 694–711, May 2006.

[13] M. Valera and S. A. Velastin, "Intelligent distributed surveillance systems: A review," Proc. Inst. Elect. Eng.—Vision, Image Signal Process, vol. 152, no. 2, pp. 192–204, Apr. 2005.

[14] W. Hu, T. Tan, and L.Wang, "A survey on visual surveillance of object motion and behaviors," IEEE Trans. Syst., Man, Cybern. C, Appl. Rev., vol. 34, no. 3, pp. 334–352, Mar. 2004.

[15] Teddy Ko, "A survey on behavior analysis in video surveillance for homeland security applications," In the proc. of 37th IEEE Applied Imagery Pattern Recognition Workshop, pp. 1-8, Oct. 2008.

[16] Arifur Rahaman, Md. Mehedi Hasan, S. M. Nafiul Hossain, M. M. Farhad, Md. Nazmul Ahassan, Mirza Md. Shahriar Maswood, and Md. Mostafizur Rahman, "An Efficient Background Updating Model for Motion Detection", In the proc. of 2nd International Conference on Informatics, Electronics & Vision (ICIEV 2013), May-17-18, 2013. (In press)

Design of a Novel Dual-Band Microstrip Patch Antenna Operating at 2.45 GHz and 2.84 GHz with Practical Implementation

Anik Mallik, Sanjoy Kundu
Department of Electronics & Communication Engineering (ECE)
Khulna University of Engineering & Technology (KUET)
Khulna, Bangladesh
anikmallik@yahoo.com, sanjoykundu.ece@gmail.com

***Abstract*—This paper presents the design and analytical studies of a novel dual-band microstrip patch antenna operating at two frequencies of ISM bands: 2.45 GHz and 2.84 GHz. The necessity for multiband antennas is increasing with the demand of wireless communications and as a result, light-weight, low-volume and low-profile microstrip patch antennas have been nourished by the researchers which initiated our design of this proposed antenna. This antenna is characterized well with its parameters i.e. Reflection coefficient (below -10 decibels), voltage standing wave ratio (VSWR) (about 1.5), forward gain (above 8.07 decibel with compared to isotropic antenna), radiation pattern etc. which are shown in this paper. The antenna was designed in Numerical Electromagnetic Code (4nec2) design environment and simulated for the antenna parameters. Then it was practically implemented and tested with WATS-2002 (wave and antenna training system). The simulated and tested results show that this antenna is eligible for short range communication like Bluetooth, WLAN, Wi-Fi and so on. This antenna is highly directive, matched with 50 ohms feeding line impedance and has uniform radiation over all the directions, but has a low bandwidth for each of the two bands which can be improved with a modification in the structure and the frequency bands can be changed by adding some parasitic elements.**

***Index Terms*—Microstrip patch, multiband antenna, ISM band, forward gain, VSWR, radiation pattern.**

I. Introduction

Modern civilization, now-a-days, completely relies on communication system advancements. The evolution of wireless communication has brought the shine of the science into the palm of human race where antenna covers the lead role in the electromagnetic paradigm. An antenna can be defined as a transducer designed to transmit or receive electromagnetic waves which converts electromagnetic waves into electric current and vice versa [1]. It is the device from which any signal is radiated into free space which makes it a very important part of any wireless systems where the 'property of reversibility' is an important feature; the same antenna can be used with the same characteristics as a transmitter or as a receiver antenna [2]. The enormous research in wireless communications has been introducing the development of new antennas with new design and measurement that operate in the required frequency bands [3]-[7]. Such as, a proximity-fed annular slot antenna for UWB applications has been designed with band rejection [8]. The design of wearable components and antennas for wireless applications has also caught significant attention in the past decade [9].

But, a small antenna with unidirectional radiation characteristics is required to be designed for the short-range wireless communication [10]. This necessitates an initiative to extensive research and development of micro-strip antenna and arrays exploiting the numerous advantages such as simple, inexpensive, light weight, low volume, low cost, planar configuration, compatibility with integrated circuits, easy to mount and mechanically robust which have led to diverse applications to the broad field of microwave antennas [11], [12]. A microstrip patch antenna offers a low profile, which means thin in structure and easy to manufacture, providing with a great advantage over traditional antennas [13]. Patch antennas are planar antenna used in wireless link communications and a wide variety of microwave applications [14]. They can be easily integrated into arrays- which is much advantageous while greater gain is needed [15]-[17].

Microstrip patch antenna is named before the radiating patch conductor it has on the board, which is generally made of copper and can be of any shape square, rectangular, circular, elliptical, triangular or any other desired configuration. Because of ease of analysis and fabrication and their attractive radiation, square, rectangular, dipole and circular are the most common in the design procedure [18]. In order to achieve the polarization, microstrip antenna usually uses two orthogonal patch modes in phase quadrature [19]. The circularly polarized microstrip patches are commonly used for various applications. Such as, a well-known circularly polarized structure, the square patch microstrip antenna with truncated corner, has been used for the UHF band [20]. Most circularly polarized patch antennas are designed on the two adjacent sides of a square patch with signals of equal magnitude and 900 phase difference using a power splitting network which can also be a truncated patch fed by single strip or coaxial line [21]. In spite of being many benefits to use microstrip patches, they have some

drawbacks, too. The operational disadvantages of micro-strip antennas are their low efficiency, low power, high Q, poor polarization purity, poor scan performance, spurious feed radiation and very narrow frequency bandwidth, which is typically only a fraction of a percent or at most a few percent, in practice [22].

Due to rapid growth of wireless communication technology, the necessities to cover multiple applications with a single antenna element have engaged researches and developments on it. Traditionally used antennas operate at single frequency bands; as a result, different antenna is needed for different applications which will eventually cause a limited space and place problem. The use of multiband antenna can rule over this problem, where a single antenna can operate at many frequency bands [23]. Antennas, which can work properly in more than one frequency bands without facing any major problem, are known as Multiband antennas [24], [25]. In the consequences, dual-frequency or dual-band antenna is one of those new antennas and so far numerous efforts have been made in this area [26]-[28]. Microstrip patches have been investigated for dual-frequency operation [29] and it is seen that such antennas are compact and have simple fabrication process, low-cost, and indeed a very versatile candidate for multiband applications [30].

Now, we should think about the next generation wireless communication systems, where at present they are heading towards high transmission speeds and high capacity via higher system frequencies to facilitate a wide frequency band and using superior system technology to attain high quality and signal multiplexing and the antenna has some prerequisites for these system properties, which are: high-gain as well as high-efficiency technology and multi-antenna technology [31]. High gain can be achieved using innovative materials such as metamaterials [32] or parasitic elements etc. On the other hand, the multi-antenna technology includes beam scanning, fixed narrow multi-beam system, MIMO antenna etc.

The tremendous growth in demand for wireless RF systems in applications such as local area networks, point-to-point communications, and applications in medical and industrial sectors exploit instrumentation, scientific and medical (ISM) bands. Inverted FL antennas are designed for realizing dual-frequency operation at 2.45 and 5.2 GHz (frequencies used for wireless LAN communications) [33]. One of the RFID systems operates at microwave frequency (2.45 GHz and 5.8 GHz) [34]. Microstrip antennas have made a perfect use in the wireless local area network (WLAN) applications, too. For high-speed broadband WLAN systems, broadband microstrip patch antennas for the 2.45 GHz ISM band are also designed [35].

The proposed antenna in this paper can be designed to be used at such frequencies. The antenna we designed has dual frequency band which lie at 2.45 GHz and 2.84 GHz. We have used the software "4nec2" (Numerical Electromagnetic Code) to design and simulate the antenna, which is actually open-source software [36]. Then we have fabricated the antenna and tested it with WATS-2002 experimental kit (wave and antenna training system) [37] and achieved the measured results. The antenna has excellent performance in both the frequencies. It has a narrow bandwidth and high gain, which make it reliable for short range communication. The frequency bands are in ISM (Industrial, Scientific and Medical) bands, where 2.45 GHz is useful for Bluetooth, WLAN, Wi-Fi, DSSS (direct sequence spread spectrum) techniques, ZigBee protocol (based on IEEE 802 standard), etc. and 2.84 GHz is for UFFS (Ultrafast Fine Structure) data receiving [38]. In the following sections, we will present the design and fabrication procedure along with relevant results and discussions.

II. Design of the Proposed Antenna

In this proposed antenna, rectangular-shaped patch is used and feeding which is used to patch antenna section to 50 Ω transmission line. Figure 1 shows the process of building the new shaped patch antenna. Here perfect ground is used to simulate this antenna. The radius of each patch is 1 mm and the height of the substrate is 1.6 mm. The antenna is designed under the 4nec2 environment.

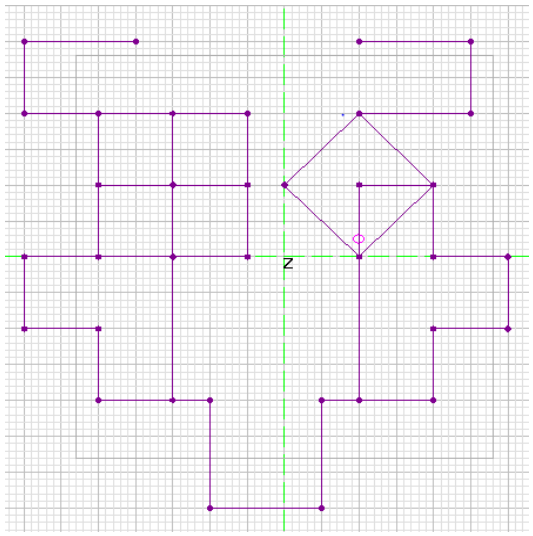


Fig. 1. Geometrical structure and feeding of the proposed antenna

III. Simulated Results

An antenna is characterized by its center frequency, bandwidth (BW), polarization, gain, radiation pattern and impedance [2]. In the following sub-sections, Reflection coefficient, VSWR, Input impedance and total gain will be examined. These results proof the applicability of this antenna as well as the novelty of the structure.

A. Reflection Coefficient

Reflection coefficient or return loss is defined as the ratio of reflected power to the incident power. For practical antenna its value should be less than −9.54 db. Simulated return loss in case of the proposed micro-strip patch antenna is −14.4 dB at

frequency 2.45 GHz and −11.3 dB at 2.84 GHz as shown in Fig. 2.

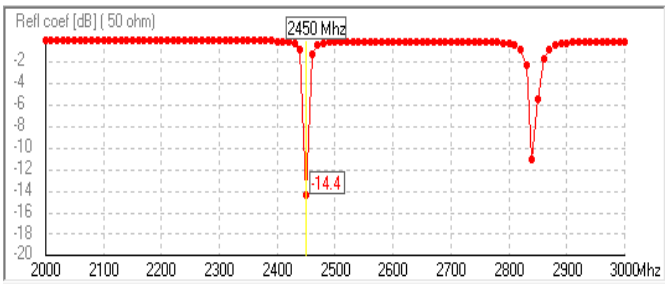


Fig. 2. Reflection coefficient Vs Frequency

B. Voltage Standing Wave Ratio (VSWR)

Voltage standing wave ratio should have the value less than 2 for the antenna to radiate. In case of designed rectangular micro-strip antenna it comes out to be 1.471 at 2.45 GHz and be 1.574 at 2.84 GHz. The VSWR curve is plotted against frequency in Fig. 3.

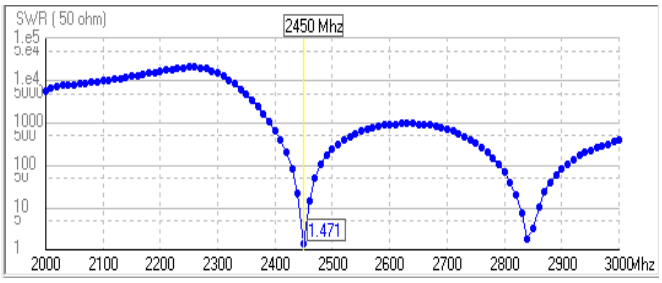


Fig. 3. VSWR Vs Frequency

C. Input Impedance

The impedance of the antenna ought to match with the feeding line impedance which is 50 Ω. It can be seen from Fig. 4 that for 2.45 GHz, the input impedance of the designed micro-strip antenna is 49.63 Ω and for 2.84 GHz, the value is 51.49 Ω.

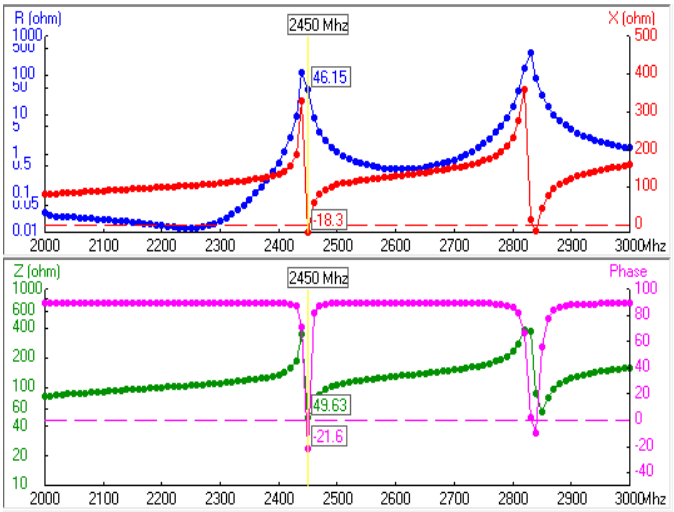


Fig. 4. Impedance Vs Frequency

D. Total Gain

The total gain curve is shown in Fig. 5 which is 8.07 dBi at 2.45 GHz and is 8.36 dBi at 2.84 GHz. And Fig. 6 depicts the gain pattern at 2.45 GHz while the antenna is steady in one direction, where dBi is a measure of the forward gain of an antenna compared with the hypothetical isotropic antenna, which uniformly distributes energy in all directions.

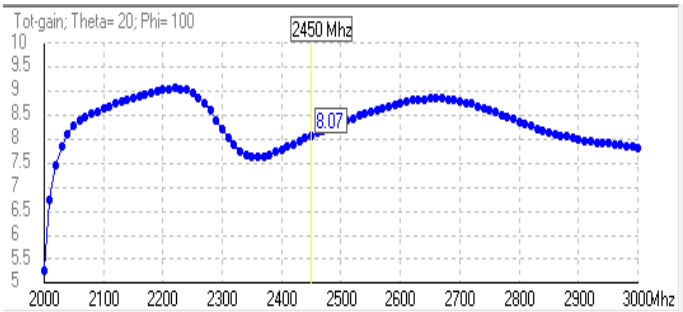


Fig. 5. Total gain Vs Frequency

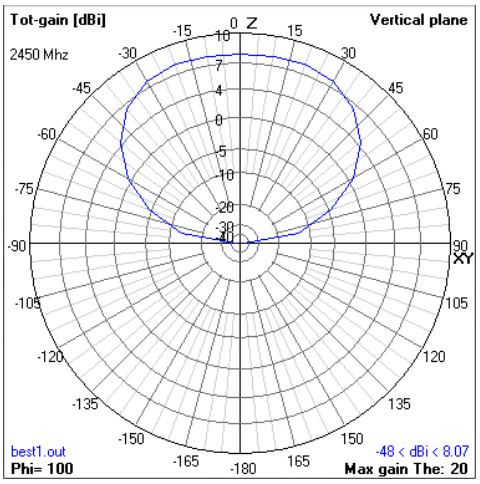


Fig. 6. Total gain pattern at 2.45GHz

IV. Practical Implementation

The proposed antenna was designed practically using PCB (Printed Circuit Boards) and some chemical compounds. At first a PCB was covered over using a transparent paper as the substrate and then the paper was cut out in accordance with the structure. Then the cut space was filled with common nail polish with accurate scaling and was then left to dry. Then the PCB substrate was drowned into a solution of Ferric Chloride ($FeCl_3$) for a while to remove the conductivity of the PCB and then it was let dry again. Then the nail polish was removed using hard paper and another PCB was attached to the back side of the modified substrate to act as ground. A connector was used to join the antenna to the feeding line. The familiar methods to feed micro-strip antenna are micro-strip line, coaxial probe, aperture coupling and proximity coupling [39], [40]. Here, a coaxial probe was used to get the required feeding. The antenna was then tested using WATS-2002. Figure 7 shows the practically implemented antenna with the patch and connector having an input impedance of 50 Ω.

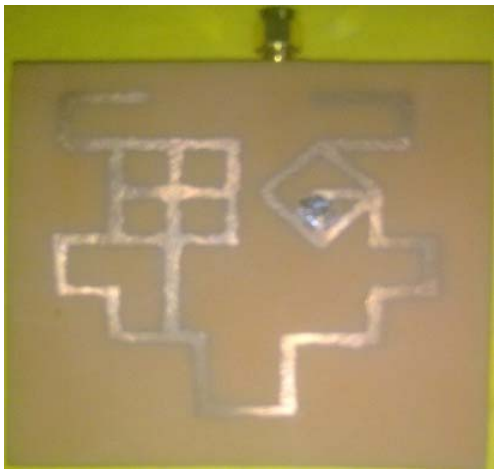


Fig. 7. Pictorial view of the practically designed antenna

V. Radiation Pattern from WATS-2002

As the thickness of micro-strip is normally very small, the waves generated within the dielectric substrate undergo reflections to some extent when they arrive at the edge of strip, resulting radiations. It is found, in practice that radiated energy from antenna is not of same strength in all directions. Instead it is more in one direction less or zero in other direction. The radiated energy in a particular direction is measured in terms of field strength at a point which at particular distance from the antenna. Radiation pattern is a graph which shows the variation of actual field strength of electromagnetic field at all points which are at equal distance from antenna. Practical antennas are designed to have directional radiation pattern, i.e. they will radiate or receive radiation more effectively in one specified direction. Our designed antenna was tested using the Wave and Antenna (WATS-2002) Training Kit and the radiation pattern was observed.

The antenna was mounted and rotated using the rotator and the radiation pattern for all the directions was achieved which is shown in Fig. 8, but only at 2.45 GHz, as WATS-2002 does not support the 2.84 GHz frequency.

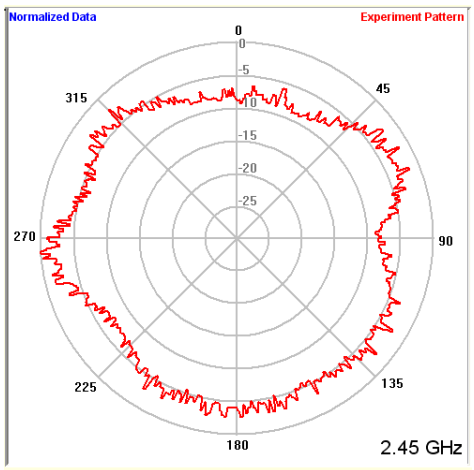


Fig. 8. Radiation Pattern form WATS-2002 at 2.45 GHz

The rotator was used to rotate the antenna about 360^0. This result from Fig. 8 simply states this antenna provides almost uniform radiation, but the radiation pattern of the designed antenna is maximum at almost 270^0. So, the antenna is capable of handling the situations where uniform radiation is required in case of short distance communication.

VI. Summarized Result & Discussion

The results for 2.45 GHz and 2.84 GHz are summarized for different characteristics parameters of the proposed antenna. Table I shows the summarized result:

TABLE I. Summarized Result of The Proposed Antenna at Dual-Bands

Parameters	At 2.45 GHz	At 2.84 GHz
Reflection coefficient	-14.4 dB	-11.3 dB
VSWR	1.471	1.574
Impedance	49.63 Ω	51.49 Ω
Total Gain	8.07 dBi	8.36 dBi

From the results, it can easily be seen that the antenna works efficiently in both the frequency bands at 2.45 GHz and 2.84 GHz. The radiation pattern is almost uniform in all the directions along 360 degrees (without any line of sight). This antenna has narrow bandwidth from 2.44 GHz to 2.46 GHz (20 MHz) and to 2.835 GHz to 2.845 GHz (10 MHz) but with intense directivity which is essential for short range communication. The antenna gain and VSWR is tremendous to be used in practical applications e. g., Bluetooth, WLAN etc.

VII. Conclusion

In the world of communication, antenna serves the greatest deal of wireless link establishment where microstrip antenna is a revolutionary step to increase directivity and to reduce size and weight which has accelerated the advancements in short range communications. Enhancement of the bandwidth and gain of Micro-strip elements still remains a challenge to the researchers. In this paper, a rectangular micro-strip patch antenna is proposed which is designed as a lightweight, low volume and low profile planar antenna in the application for Bluetooth, WLAN, Wi-Fi, DSSS systems and many other applications, at a frequency 2.45 GHz and 2.84 GHz. The gain and VSWR of the antenna are so satisfactory that they can be fabricated commercially to a greater extent. This dual-band microstrip patch antenna is simulated and tested practically which gave almost the same results. This device has a well directivity as well as well radiation pattern, but a low bandwidth. By modifying the substrate height or structural parameters, the bandwidth can be enhanced, which is a concern of ours. Again, this antenna works in the ISM band only, but adding some parasitic elements will change the center frequency and thus change the operating band. We have a plan to enhance this antenna's bandwidth in future as well as its gain which, we think, can be done by adding some parasitic elements such as more patches, LC networks etc. We also wish to make the antenna to provide more operating bands.

ACKNOWLEDGMENT

We want to acknowledge gratefully Mr. Md. Mehedi Hasan, a friend of ours, for helping us to turn this research work into a well-written form. Special gratitude goes to Mr. Khaled Mahbub Morshed (Assistant Professor) and Prof. Dr. Md. Mostafizur Rahman from Department of Electronics & Communication Engineering, Khulna University of Engineering & Technology, for their valuable suggestions and also for providing the facilities to test the antenna with WATS-2002.

REFERENCES

[1] C. A. Balanis, Antenna theory: analysis and design, Wiley, New York, 1996.

[2] John D. Kraus, and Ronald J. Marhefka, Antennas for all Applications, 3rd Edition, McGraw-Hill, 2002.

[3] Y. Guo, M. Y. Chia, and Z. N. Chen, "Miniature built-in quad-band antennas for mobile handsets", IEEE Antennas Wireless Propag. Lett., vol. 2, pp. 30-32, 2003.

[4] M. Ali, G. J. Hayes, H. Hwang, and R. A. Sadler, "Design of multiband internal antenna for third generation mobile phone handsets", IEEE trans. Antenna Propag., vol. 51, no. 7, pp. 1452-1461, Jul 2003.

[5] Y. Guo, I. Ang, and M. Y. W. Chia, "Compact internal multiband antennas for mobile handsets", IEEE Antennas Wireless Propag. Lett., vol. 2, pp. 143-146, 2003.

[6] M. F. Abedin, and M. Ali, "Modifying the ground plane and its effect on planar inverted-F antennas (PIFAs) for mobile phone handsets", IEEE Antennas Wireless Propag. Lett., vol. 2, pp. 226-229, 2003.

[7] Y. Kwon, J. Moon, and S. Park, "An internal triple-band planar inverted-F antenna", IEEE Antennas Wireless Propag. Lett., vol. 2, pp. 341-344, 2003.

[8] E. E. M. Khaled, A. A. R. Saad, and D. A. Salem, "A Proximity-fed annular slot antenna with different A Band-notch manipulations for ultra-wide band applications", Progress In Electromagnetics Research B, vol. 37, pp. 289-306, 2012.

[9] Hertleer, C., Rogier, H., Vallozzi, L., and Van Langenhove, L., "A Textile Antenna for Off-Body Communication Integrated Into Protective Clothing for Firefighters", IEEE Trans. Antennas Propag., vol. 57, no. 4, pp. 919-925, 2009.

[10] Mitsuo Taguchi, and Shinya Tanaka, "Two element phased array dipole antenna on EBG ground plane", In the Proceedings of ISAP 2007, Niigata, Japan.

[11] R. Garg, P. Bhartia, I. Bahl, and A. Ittipiboon, Microstrip Antenna Design Handbook, Artech House Antennas & Propagation Library, Nov 2000.

[12] Vandana Vikas Thakare, Shailendra Singh Jadon, and Ritesh Kumari, "Analysis of Circular Microstrip Patch Antenna using Radial Basis Artificial Neural Network", International Journal of Electronics and Computer Science Engineering, vol. 1, no. 3., 2012.

[13] "Microstrip patch antenna," www.electronicshome.com, Jan. 2007.

[14] Rajeshwar Lal Dua, Himanshu Singh, and Neha Gambhir, "2.45 GHz Microstrip Patch Antenna with Defected Ground Structure for Bluetooth", International Journal of Soft Computing and Engineering (IJSCE), vol. 1, issue 6, January 2012.

[15] J. T. Abrele, and D. M. Pozar, "Analysis of infinite arrays of one and two probe fed circular patches," IEEE Transactions Antenna and Wave Propagation, vol. 38, no. 4, pp. 991-994, Sept. 1984.

[16] Telikepalli, R., Strickland, P.C., McKay, K.R., and Wight, J.S., "Wide band microstrip phased array for mobile satellite communications," IEEE Transactions Microwave Theory Tech. vol. 43, no. 7, pp. 1758-1763, July 1995.

[17] Jedlicka, R. P., Poe, M., and Carver, K., "Measured mutual coupling between microstrip antennas," IEEE Transactions Antenna and Wave Propagation, vol. 29, no. 1, pp. 147-149, Jan. 1981.

[18] K. D. Parsad, Antenna and Wave Propagation, Satya Parkashan, 2005.

[19] H. F. Lee, and W. Chen, Advances in Microstrip and Printed Antennas, Wiley, New York, 1997.

[20] J. R. James, and P. S. Hall, Handbook of Microstrip Antennas, Peter Peregrinus Ltd., London, U.K., 1989.

[21] Xianming Qing, and Ning Yang, "2.45 GHz Circularly polarized RFID reader antenna", ICCS, 2004.

[22] A. P. Singh, and J. Singh, "On the Design of a Rectangular Microsrip Antenna Using ANN", National Journal of The Institution Of Engineers (India), vol. 29, pp. 20-25, July 2009.

[23] Amit K. Panda, and Asit K. Panda, "A Novel Design of Multiband Square Patch antenna embedded with Gasket Fractal Slot for WLAN & WIMAX communication", International Journal of Advances in Engineering & Technology, March 2012.

[24] G. R. DeJean, Compact Broadband and Multiband Antenna Designs for Millimeter-Wave Applications, VDM Verlag Dr. Mueller e.K., Saarbruecken, Germany, 2008.

[25] G. Kumar, and K. P. Ray, Broadband Microstrip Antenna, Artech House, Norwood, MA, 2003.

[26] D. Llorens, P. Otero, and C. C-Penalosa, "Dual-band, single CPW port, planar-slot antenna", IEEE Trans. Antennas Propag., vol. 51, no. 1, pp. 137-139, Jan. 2003.

[27] S. Yeh, K Wong, T. Chiou, and S. Fang, "Dual-band planar inverted-F antenna for GSM/DCS mobile phones", IEEE Trans. Antennas Propag., vol. 51, no. 5, pp. 1124-1126, May 2003.

[28] H. Chen, W. Chen, Y. Cheng, and Y. Lin, "Dualband meander monopole antenna", in the Proc. IEEE Antennas and Propagation Soc. Int. Symp., vol. 3, pp. 48-51, Columbus, OH, June 2003.

[29] K. Wong, Compact and Broadband Microstrip Antennas, New York, Wiley 2002.

[30] Rakhesh Singh Kshetrimayum, "Printed Double-T monopole antennas for Triband applications", Microwave and Optical Technology Letters, vol. 51, no. 7, July 2009.

[31] Pratik Bhanawat, "Current Antenna System Technologies for Next-Generation Wireless Communication", 16th International Conference on Network-Based Information Systems (NBiS 2013), in press.

[32] Anik Mallik, Sanjoy Kundu and Md. Osman Goni, "Design of a Novel Two-Rectangular U-Shaped Double Negative Metamaterial", in the proceedings of 2nd International Conference on Informatics, Electronics & Vision (ICIEV) 2013, Dhaka University, Bangladesh.

[33] Hisamatsu Nakano, Yusuke Sato, Hiroaki Mimaki, and Junji Yamauchi, "An Inverted FL Antenna for Dual-Frquency Operation", IEEE Transactions on Antennas and Propagation, vol. 53, no. 8, pp. 2417-2421, August 2005.

[34] K. Finkenzeller, RFID Handbook, 2nd Ed., John Wiley & Sons, Chichester, 2004.

[35] Salman Haider, and Lindsay Young, "Microstrip Patch Antennas for Broadband Indoor Wireless Systems", Project Report, Supervisor: Dr. Michael Neve, University of Auckland.

[36] 4nec2, NEC based antenna modeler and optimizer, website: http://www.qsl.net/4nec2.

[37] WATS-2002, Wave and Antenna Training System, website: http://eng.manntel.com.

[38] Zheng Le-Ping, and Liu Yu-Ying, "A 2.84 GHz MS-resolution receiver and the question about the reliability of the 205 Solar UFFS data taken with it", Solar Physics 146, pp. 191-197, Kluwer Academic Publishers, Belgium, 1993.

[39] I. J. Bahl, and P. Bhartia, Microstrip Antennas, Artech House, Dedham, MA, 1980.

[40] K. R. Carver, and J. W. Mink, "Microstrip Antenna Technology", IEEE Transactions on Antenna and wave propagations, vol. 29, pp. 2-24, Jan.1981.

Performance of Turbo coded Vertical Bell Laboratories Layered Space Time Multiple Input Multiple Output System

M.M.Kamruzzaman
Key Lab of Information Coding & Transmission,
Southwest Jiaotong University, Chengdu, Sichuan, China.
E-mail: m.m.kamruzzaman@gmail.com

Abstract— **In this paper, performance of a Turbo Code (TC) and Vertical Bell Laboratories Layered Space Time (VBLAST) encoded wireless link is evaluated in the presence of Rayleigh fading for Multiple-Input Multiple-Output (MIMO) system. Turbo encoded bits are modulated using QPSK or 16 QAM or 64 QAM modulator and modulated data are further encoded using VBLAST and then split into *n* streams which are simultaneously transmitted through *n* transmit antennas. Simulation results obtained show that coded MIMO system provides 17 to 25 dB coding gain for different combination of transmit and receive antennas at a BER of 10^{-3} compared to uncoded MIMO system. And there is around 2 to 8 dB gain for increasing number of transmitting antennas from 2 to 3 or 3 to 4. On the other hand, capacities for ($2\ T_x$, $2\ R_x$), ($3\ T_x$, $3\ R_x$) and ($4\ T_x$, $4\ R_x$) are almost same from 0dB to 10dB but change significantly with increasing SNR from 10dB.**

Keywords—VBLSAT; MIMO; Turbo Code;

I. INTRODUCTION

As the demand grows for capacity in wireless communication under limited radio spectrum, Multiple Input Multiple Output (MIMO) technologies are becoming increasingly important. There are many MIMO techniques which have been proposed based on the different scenarios in wireless communication. On of the techniques which are promised for huge capacities in fading environment is Spatial Multiplexing (SM). In SM, Information are divided into sub-streams and multiple transmit antennas are used to transmit these sub-streams information through several independent (spatial) channel simultaneously on the same carrier frequency. The same frequency bandwidth is used and the total transmission power is always held constant. On the other hand, multiple receiving antennas are used to receive the signal. So the data transmission rate is directly proportional to the number of antennas. The higher the number of antennas, the higher the data transmission rate.

Vertical-Bell-Labs-Layered-Space-Time-Architecture (VBLAST) is an example for SM introduced by Gerard J. Foscini in 1996 at Bell Laboratories [1]. It has been shown that the theoretical capacity approximately increases linearly as the number of antennas is increased.

However, only VBLAST technique can't satisfy the reliability requirement in future mobile systems, so it should be concatenated with channel coding to provide more coding gains. Forward Error correction (FEC) coding schemes are used as channel coding in most of the digital communication systems. Turbo Codes (TC) are a class of high-performance FEC codes which were the first practical codes to closely approach the channel for the SISO system capacity[2-4] which are specified as FEC schemes for most of the wireless systems. A concatenation of the turbo coding and VBLAST referred to as the Turbo-VBLAST.

We published several papers from our previous research on Turbo coded Space time Block Code (STBC) with and without OFDM [5-11]. We also published a paper on Turbo VBLAST coded OFDM system [12]. In this paper, we have studied turbo coded VBLAST system without OFDM. There are several papers on turbo coded VBLAST system [13-20] but our scheme performs better than earlier proposed schemes with respect to BER.

II. SYSTEM MODEL

The system considered in this paper consists of n transmit and m receive antennas as shown in fig1 and fig 2.The Encoding and Decoding techniques of the system are discussed in detail below.

A. Encoding

Data are encoded by a turbo encoder and the encoded bits are modulated by a QPSK or 16 QAM or 64 QAM modulator and the modulated symbols are demultiplexed into n separate streams, using a serial-to-parallel converter, and each stream is transmitted from an independent transmit antenna. Suppose at time *t*, signals s_t^i , i =1, 2, ... n are transmitted simultaneously using n transmit antennas. The transmission sequences are shown in Table I, Table II and Table III for different number of transmit antennas.

B. Channel

We assume that the transmitted signals encounter flat fading channel before reaching the receiving antennas. The path gain from transmit antenna i to receive antenna j is denoted by $\alpha_{i,j}$. The path gains are modeled as samples of independent and identically distributed (i.i.d), zero mean, complex Gaussian, unit variance entries (the variance of each

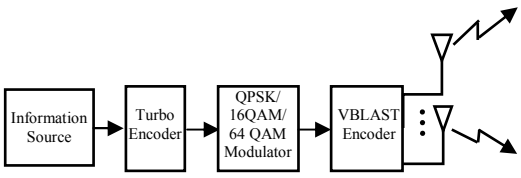


Fig 1. Block diagram of transmitter

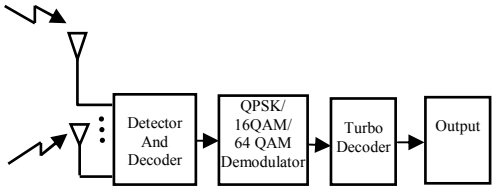


Fig 2. Block diagram of receiver

entry is $\sigma_c^2 = 1$) .The wireless channel is assumed to be quasi-static so that the path gains are constant over a frame of length l and vary from one frame to another.

C. Received Signal

At time t the signal y_t^j , received at antenna j of destination, is given by

$$y_t^j = \sum_{i=1}^{n} \alpha_{i,j} s_t^i + \eta_t^j \tag{1}$$

where y_t^j is the $m \times 1$ dimention complex vector of received symbols.

$\alpha_{i,j}$ is the $m \times n$ dimention channel matrix.

s_t^i is the $n \times 1$ dimention complex vector of transmitted sub-streams, each assumed to have zero-mean, uncorrelated random variables with variance equal to σ_s^2 .

TABLE I. THE TRANSMISSION SEQUENCE FOR TWO TRANSMISSION ANTENNAS

	Antenna-I/ Layer-I	Antenna-II/ Layer-II
Time slot-I	x_1	x_2
Time slot-II	x_3	x_4

TABLE II. THE TRANSMISSION SEQUENCE FOR THREE TRANSMISSION ANTENNAS

	Antenna-I/ Layer-I	Antenna-II/ Layer-II	Antenna-III /Layer-IIII
Time slot-I	x_1	x_2	x_3
Time slot-II	x_4	x_5	x_6
Time slot-III	x_7	x_8	x_9

TABLE III. THE TRANSMISSION SEQUENCE FOR FOUR TRANSMISSION ANTENNAS

	Antenna-I/ Layer-I	Antenna-II /Layer-II	Antenna-III/ Layer-III	Antenna-IV/ Layer-IV
Time slot-I	x_1	x_2	x_3	x_4
Time slot-II	x_5	x_6	x_7	x_8
Time slot-II	x_9	x_{10}	x_{11}	x_{12}
Time slot-IV	x_{13}	x_{14}	x_{15}	x_{16}

η_t^j is is the $m \times 1$ dimention complex vector representing the additive receiver noise, assumed to have zero-mean, uncorrelatd random variables with variance σ_n^2 .

D. Decoding

The receiver can be configured by different types of detection techniques such as ZF, MMSE and ML for VBLAST. In this paper, we considered ZF and ML to show the performance of the system.

a) ZF

In this technique, each sub-stream in turn is considered to be the desired signal and the remaining data streams are considered as "interferers". Nulling of the interferers is performed by linearly weighting the received signals sothat all interfering terms are cancelled. Main steps are as follows [21]:

Step 1: Ordering - The purpose of the ordering is to decide which transmitted symbol to detect at each stage of the decoding. The symbols with highest SNR are selected first and then select the symbols in the order of decreasing SNR.

Step 2: Interference cancellation - The goal of the interference cancelation is to remove the interference from the already detected symbols. i.e. subtract the detected symbol $\hat{s}$ from the received vector y .

Consider At stage l of the algorithm, when s_l is being detected, symbols $s_1, s_2, \ldots, s_{l-1}$ have been already detected. Let us assume a perfect decoder, that is the decoded symbols $\hat{s}_1$, $\hat{s}_2$,…, $\hat{s}_{l-1}$ are the same as the transmitted symbols $\hat{s}_1, \hat{s}_2, \ldots, \hat{s}_{l-1}$. One can subtract $\sum_{i=1}^{l-1} s_i \alpha_i$ from the received vector y to derive an equation that relates remaining undetected symbols to the received vector:

$$y_l = y - \sum_{i=1}^{l-1} s_i \alpha_i$$

$$y_l = \sum_{i=1}^{l-1} s_i \alpha_i + \eta \qquad l = 2,3,\ldots,L-1. \tag{2}$$

In fact, by using induction in addition to the convention $y_l = y$, one can show that

$$y_{l+1} = y_l - s_l \alpha_l, \; n = 1,2,\ldots,L-2. \tag{3}$$

Therefore, at the lth stage of the algorithm after detecting the lth symbol as $\hat{s}_l$, its effect is canceled from the equation by

$$y_{l+1} = y_l - \hat{s}_l \alpha_l \quad (4)$$

Step 3: Interference Nulling - Interference Nulling is the process of detecting s_l from y_l by first removing the effects of undetected symbols. Basically, in this step the lth symbol is detected by nulling the interference caused by symbols $s_{l+1}, s_{l+2}, \ldots, s_L$.

So, we would like to separate the $s_l \alpha_l$ from y_l. This can be done through multiplying y_l by an $m \times 1$ vector W_l that is orthogonal to interference vectors $\alpha_{l+1}, \alpha_{l+2}, \ldots, \alpha_L$ but not orthogonal to α_l. In other words, W_l should be such that

$$\alpha_i W_l = 0, \quad i = l+1, l+2, \ldots, L,$$
$$\alpha_l W_l = 1. \quad (5)$$

In fact, W_l is called the zero-forcing nulling vector with minimum norm. Such a cevtor is uniquely calculated from the channel matrix α. To calculate W_l from α, for $m \geq n$, first we should replace the rows $1,2,\ldots,l-1$ of α by zero. Let us denote the resulting matrix by Z. Then, W_l is the lth column of Z^+, the Moore-Penrose generalized inverse, pseudo-inverse, of Z.

Using the error free detection formula for y_l in (2) and W_l in (5), we have

$$y_l . W_l = s_l + \eta . W_l \quad (6)$$

The noise in (6) is still Gaussian and the symbols s_l can be easily decoded.

Step 4: Slicing - Making a symbol decision. i.e. s_l can be sliced to the nearest QAM constellation point, these sliced signals are denoted by $\hat{s}$. The decoded symbol $\hat{s}_l$ is the closest constellation point to $y_l . W_l$.

Step 5: Iteration - Going to the first step to detect the next symbol

b) ML:

ML is a method that compares the received signal with all possible transmitted signals and estimates s according to the maximum Likelihood principle. Suppose a matrix C gives all possibilities in s that could occur (the dimensions of C are $n \times K$ where $K = M^n$ and M represents the number of constellation points) Then, the receiver should store a matrix Z such that:

$$Z = \alpha . C = [z_1 \ldots z_k] \quad (7)$$

At the receiver, the most likely transmitted signal is determined, as the one for which $\|y - z_j\|^2$ is minimal (with $1 \leq j \leq k$, i.e., the signal s_j that corresponds with the vector z_j which lays closed to he received vector is said to be the most likely signal to be transmitted. Thus $\hat{s}$ is chosen to be the jth column of C .This can be rewritten to the following formula where s_{ml} represents the maximum likelihood detection of the transmitted signal s:

$$\hat{s} = s_{ml} = \arg \min_{s_j \in \{s_1 \ldots s_k\}} \|y - \alpha s_j\|^2 \quad (8)$$

Note that in the case of ML, it is not required that $n \leq m$. The detected symbols are demodulated by 64QAM demodulator and send to turbo decoder to get the output.

E. Capacity

The SNR for each component of y, can be found and is equal to:

$$SNR = \frac{E_s}{N_0} = \frac{n\sigma_s^2 \sigma_c^2}{\sigma_n^2} = \frac{n\sigma_s^2 .1}{\sigma_n^2} = \frac{n\sigma_s^2}{\sigma_n^2} = \rho$$

Where E_s stands for the signal power per receive antenna and N_0 denotes the noise power per receive antenna.

The open-loop Shannon capacity of the MIMO system with $n \times m$ antennas, in terms of spectral effiency i.e bits per second per hz, is given by [1]

$$C = \log_2 \det\left(I_{n_R} + \frac{\rho}{n} HH^H \right) \quad (9)$$

Where, "det" means determinant, I_m is the $m \times m$ identity matrix and HH^H is an $m \times m$ positive semi-definite Hermitian matrix. " ^{H} " means transpose conjugate.

It is assumed for (9) that the transmitter is constrained to communicate using i.i.d random process of equal power from each of n antennas. It is to mention that the capacity in (9) can be only achieved with the use of strong encoding (STC) techniques.

If $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_{\min}$ are the (random) ordered singular values of H, then we can rewrite () as [22]

$$C = \sum_{t=1}^{n_{\min}} \log_2 (1 + \frac{\rho}{n} \lambda_t^2) \quad (10)$$

Where $n_{\min} = \min(n, m)$

and $\|H\|^2 = \sum_{i=1}^{m} \lambda_i$ where λ_i $(i = 1,2,...,m)$ are the eigenvalues of HH^H

The squared Frobenius norm of H, as [23] $\|H\|^2 = Tr\left|(HH^H)\right| = \sum_{i=1}^{m} \sum_{j=1}^{n} \left|\alpha_{i,j}\right|^2$. Frobenius norm is interpreted as the total power gain of the channel.

So, the Shannon capacity can be calculated for a system with n transmit antennas and m receive antennas(MIMO) as:

$$C = \sum_{t=1}^{n_{\min}} \log_2\left(1 + \left(\frac{\rho}{n}\right) \sum_{i=1}^{m} \sum_{j=1}^{n} \left|\alpha_{i,j}\right|^2\right) \qquad (11)$$

III. SIMULATION RESULTS

In this section, computer simulation is carried out to show the BER performance of the proposed system. The results are evaluated for several combinations of T_x and R_x antennas with and without Turbo coding. For the uncoded system (without turbo code), we used only VBLAST and for the coded system we used turbo code and VBLAST .

Turbo code with frame size= 378, rate= 1/3, encoder generator g = [1 0 1 1; 1 1 0 1; 1 1 1 1] and number of iterations =2 is considered to perform simulation. We present the BERs to compare the performance of coded MIMO system with uncoded MIMO system in Fig. 3. It is observed that the coded MIMO system with two T_x antennas provides 25, 19 and 17 dB coding gain for 2,3 and 4 R_x antennas respectively compared to uncoded MIMO system at a BER of 10^{-3}. There is around 3-7 dB gain for increasing Rx antennas from 2 to 3 and 3 to 4 of coded MIMO system.

Fig. 4 shows the performance of Turbo-MIMO system with 3 T_x antennas. It provides 23, 18 and 17 dB coding gain for 3, 4 and 5 R_x antenna respectively at BER of 10^{-3} compared to uncoded MIMO system. There is around 2-8 dB gain for increasing Rx antenna from 3 to 4 and 4 to 5 of coded MIMO system.

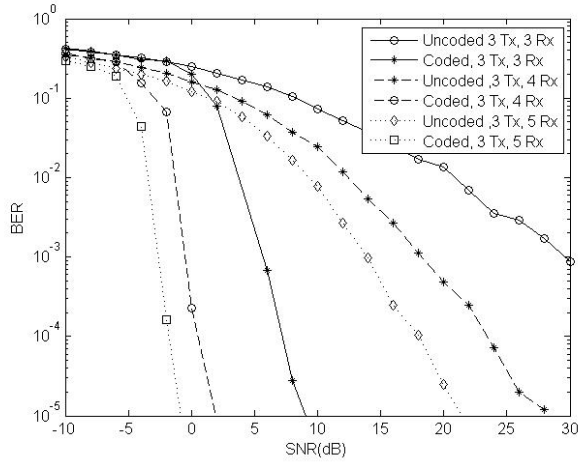


Fig. 4. BER performance comparison of coded MIMO system and uncoded MIMO system for $3T_x$ & 3/4/5 R_x .

Fig. 5 shows the performance of Turbo-MIMO system with 4 T_x antennas. It provides 24, 18 and 17 dB coding gain for 4,5 and 6 R_x antenna respectively at BER of 10^{-3} compared to uncoded MIMO system. There is around 3-5 dB gain for increasing Rx antenna from 3 to 4 and 4 to 5 of coded MIMO system with this configuration.

Fig. 6 shows the capacities of (2 T_x and 2 R_x), (3 T_x and 3 R_x) and (4 T_x and 4 R_x) system. It is interesting to note that all configuration achieve same capacity from 0-10 dB, however the gap gets increasingly larger as SNR grows from 10dB- 30 dB.

IV. CONCLUSION

From the simulations results, it is observed that TC-VBLAST encoded MIMO system makes a significant difference over uncoded VBLAST encoded MIMO system. It is possible to get 17 - 25 dB gain for different combination of transmit and receive antennas at a BER of 10^{-3} with TC-VBLAST encoded MIMO system. And there is around 2 to 8 dB gain for increasing number of transmitting antennas from 2 to 3 or 3 to 4.

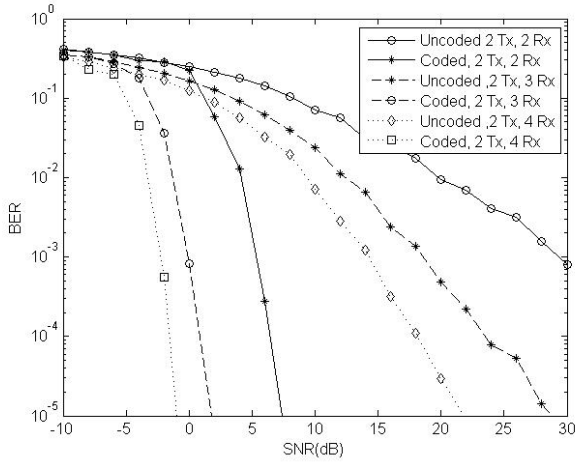


Fig. 3. BER performance comparison of coded MIMO system and uncoded MIMO system for $2T_x$ & 2/3/4 R_x .

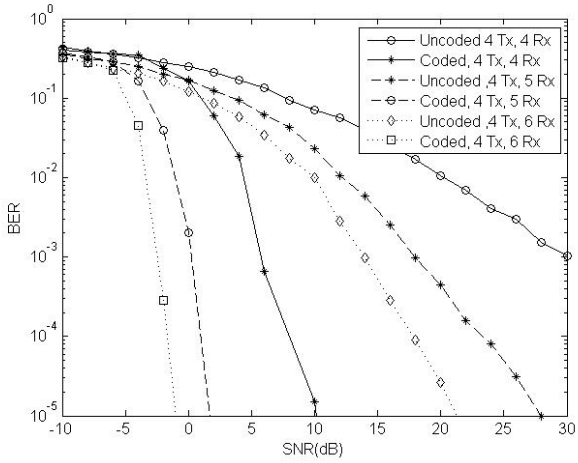


Fig. 5. BER performance comparison of coded MIMO system and uncoded MIMO system for $4T_x$ & 4/5/6 R_x .

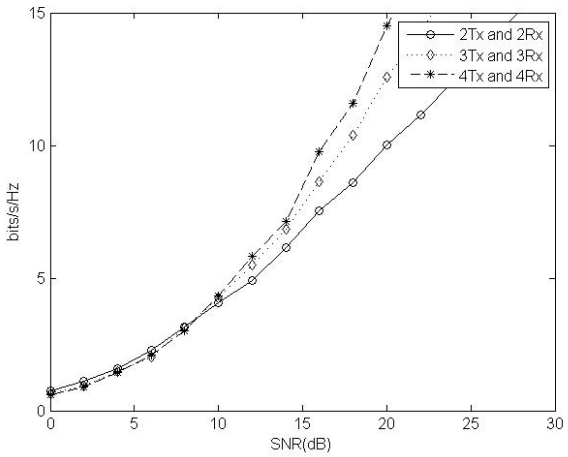


Fig. 6. Capacity for 2x2, 3x3 and 4x4 MIMO system

ACKNOWLEDGMENT

The authors would like to thank the reviewers for the suggestions which help to improve the quality of this paper. In addition, the authors are also very thankful to Key Lab of Information Coding & Transmission, Southwest Jiaotong University, Chengdu, Sichuan, China for providing resources.

REFERENCES

[1] Foschini, Gerard J. "Layered space - time architecture for wireless communication in a fading environment when using multi - element antennas."*Bell labs technical journal* 1.2 (1996): 41-59.

[2] C. Berrou, A. Glavieux and P. Thitimajshima, “Near Shannon limit error-correcting codingand decoding:Turbo-codes (1),” ICC'93. 1993: l064-1074P

[3] C. Berrou and A. Giavieux. “Near optimum error correcting coding and decoding: Turbo-codes,” IEEE Trans. Commun., vol. 44, no. 10. Oct., 1996:1261-1271P

[4] C. Berrou and A.Giavieux. Reflections on the prize paper: “Near optimum error correcting coding and decoding: turbo-codes,” IEEE Inform. Theory Society Newsletter, vol.48, no.2. June,1998: 23-31P

[5] Kamruzzaman, M.M., "Performance of Turbo Coded Wireless Link for SISO, SIMO, MISO and MIMO System," Multimedia Information Networking and Security (MINES), 2012 Fourth International Conference on , vol., no., pp.187,190, 2-4 Nov. 2012.

[6] ZAMAN, M., HAO, L.. Performance of Turbo-SISO, Turbo-SIMO, Turbo-MISO and Turbo-MIMO system using STBC. Journal of Communications, North America, 6, nov. 2011

[7] Kamruzzaman, M.M., "Performance of Turbo coded wireless link for SISO-OFDM, SIMO-OFDM, MISO-OFDM and MIMO-OFDM system," Computer and Information Technology (ICCIT), 2011 14th International Conference on, vol., no., pp.185-190, 22.

[8] Kamruzzaman, M.M., " Performance of decode and forward MIMO relaying for wireless uplink," Computer and Information Technology (ICCIT), 2012 15th International Conference on , vol., no., pp.321,325, 22-24 Dec. 2012.

[9] KAMRUZZAMAN, M., LI, H.. Performance of Turbo coded OFDM wireless link for SISO, SIMO, MISO and MIMO system. Journal of Communications**,** North America, 7, nov. 2012.

[10] Kamruzzaman, M.M., "Performance of Turbo Coded Wireless Link for SIMO Using SC EGC and MRC," Multimedia Information Networking and Security (MINES), 2012 Fourth International Conference on , vol., no., pp.191,194, 2-4 Nov. 2012.

[11] M.M.Kamruzzaman, Dr. Mir Mohammad Azad, “Single Input Multiple Output (SIMO) Wireless Link with Turbo Coding”, International Journal of Advanced Computer Science and Applications, Vol. 1, No. 5, pp. 69-73, November 2010.

[12] Kamruzzaman, M. M.; Hao, Li, “Performance of Turbo-VBLAST Coded Wireless Link for Multiple-Input Multiple-Output-Orthogonal Frequency Division Multiplexing System” Sensor Letters, Volume 10, Number 8, December 2012 , pp. 1911-1917(7)

[13] Yeong-Luh Ueng; Chia-Jung Yeh; Mao-Chao Lin; Chung-Li Wang, "Turbo coded multiple-antenna systems for near-capacity performance," Selected Areas in Communications, IEEE Journal on , vol.27, no.6, pp.954,964, August 2009.

[14] ten Brink, S.; Kramer, G., "Design of repeat-accumulate codes for iterative detection and decoding," Signal Processing, IEEE Transactions on , vol.51, no.11, pp.2764,2772, Nov 2003.

[15] van Zelst, A.; Van Nee, R.; Awater, G.A., "Turbo-BLAST and its performance," Vehicular Technology Conference, 2001. VTC 2001 Spring. IEEE VTS 53rd , vol.2, no., pp.1282,1286 vol.2, 2001.

[16] Guosen Yue; Xiaodong Wang, "Optimization of irregular repeat accumulate codes for MIMO systems with iterative receivers," Wireless Communications, IEEE Transactions on , vol.4, no.6, pp.2843,2855, Nov. 2005.

[17] Hochwald, B.M.; Ten Brink, S., "Achieving near-capacity on a multiple-antenna channel," Communications, IEEE Transactions on , vol.51, no.3, pp.389,399, March 2003.

[18] Sellathurai, M.; Haykin, Simon, "TURBO-BLAST for high-speed wireless communications," Wireless Communications and Networking Confernce, 2000. WCNC. 2000 IEEE , vol.1, no., pp.315,320 vol.1, 2000.

[19] Ten Brink, S.; Hochwald, B.M., "Detection thresholds of iterative MIMO processing," Information Theory, 2002. Proceedings. 2002 IEEE International Symposium on , vol., no., pp.22,, 2002.

[20] Stefanov, A.; Duman, T.M., "Turbo-coded modulation for systems with transmit and receive antenna diversity over block fading channels: system model, decoding approaches, and practical considerations," Selected Areas in Communications, IEEE Journal on , vol.19, no.5, pp.958,968, May 2001.

[21] Hamid Jafarkhani, Space-Time Coding: Theory and Practice, Cambridge University Press, New York, NY, 2010

[22] David Tse, Berkeley, Pramod Viswanath, Fundamentals of Wireless Communication, September 10, 2004.

[23] Mohinder Jankiraman, Space time codes and MIMO systems, Norwood, MA: Artech House, 2004.

Design and Performance Analysis of a 4-element Planar Inverted-F Antenna (PIFA) Array including Human Interaction

Md. Ashikur Rahman, Amit Thakur, and Md. Osman Goni
Department of Electronics and Communication Engineering
Khulna University of Engineering and Technology
Khulna, Bangladesh
ashik_ece_kuet@yahoo.com, amit_kuet2k8@yahoo.com, osmanhku@gmail.com

***Abstract*—A 4-elements PIFA array antenna is designed for the laptop Wi-Fi application. The proposed array antenna elements are optimized for 2.45 GHz. It is a wide band antenna. It is embedded on the back side of the laptop screen. These array elements are excited simultaneously. The performance of this antenna is analyzed and the interaction between this antenna array and user (phantom) of the laptop is studied. The VSWR and the return loss values are excellent. The SAR value lies within the acceptable limit of Federal Communications Commission (FCC). All the simulation results are performed using CST Microwave Studio 2011.**

***Keywords*—SAM Phantom; MIMO Antenna; Specific Absorption Rate (SAR); CST Microwave Studio;**

I. INTRODUCTION

Now-a-days, it is a great consideration of human body interaction with antennas of various electronic devices. Voltage standing wave ratio (VSWR), gain and radiation patterns of the antenna of the devices are influenced by the human body. Unlimited electromagnetic energy can cause a serious health hazard when tissues are exposed to electromagnetic radiation. In this case, specific absorption rate (SAR) has been termed by standard organizations to set exposure limits [1, 2]. The overall laptop integration effects on the performance of commercial 2.4 GHz Inverted-F antennas built into PCMCIA cards is studied [3]. The effects of the electromagnetic antenna-human interaction for a laptop MIMO antenna system are studied [4]. PIFAs are conducting plates called radiating patches. A shorted patch antenna with an L-shaped ground plate for internal mobile hand-set antenna is designed and analyzed in [5]. Quasi-omnidirectional dual-band back-to-back E-shaped patch antenna for laptop applications is designed and analyzed in [6]. L-bent omnidirectional patch antenna for wireless applications in laptop computers is designed and analyzed in [7]. Multi-element patch antenna integration into laptops for multi-standard applications is analyzed in [8]. Double L - slot microstrip patch antenna array for Wi-MAX and WLAN is designed and analyzed in [9]. Very-wideband microstrip antenna for laptop applications is designed in [10].

II. METHODOLOGY

A. CST Microwave Studio

CST Microwave Studio™ (CST MWS) is a specialist tool for the 3D EM simulation of high frequency components. It enables the fast and accurate analysis of high frequency (HF) devices such as antennas.

B. S-Parameter

In antenna designing s-parameter is an important parameter. It is also called scattering parameter. S-parameters describe the input-output relationship between ports (or terminals) in an electrical system. For instance, if we have two ports (intelligently called port 1 and port 2), then S_{12} represents the power transferred from port 2 to port 1. S_{21} represents the power transferred from port 1 to port 2.Specific Absorption Rate (SAR)The electromagnetic fields caused by currents on antenna surface and edges, propagates where the head and hand of user is nearby. As a result, cell polarization in human body happens due to heat induced in close body parts. Such exposure, if not controlled, can damage user health. This power absorbed by user is measured by a parameter called specific absorption rate, SAR, and is defined as the following equation.

$$SAR = \sigma \frac{|E_t|^2}{\rho} = \frac{J^2}{\sigma\rho} \quad (1)$$

At which point SAR is in (W/kg). Also σ is the conductivity and ρ is the mass density, E_t is the total amplitude of electrical field in root mean square. An average integral is used locally over a mass block to transfer from point SAR to average SAR.

C. Voltage Standing Wave Ratio (VSWR)

The voltage standing wave ratio (VSWR) at a port, represented by the lower case 's', is a similar measure of port match to return loss but is a scalar linear quantity, the ratio of the standing wave maximum voltage to the standing wave minimum voltage. It therefore relates to the magnitude of the voltage reflection coefficient and hence to the magnitude of either S_{11} for the input port or S_{22} for the output port.

At the input port, the VSWR (s_{in}) is given by

$$s_{in} = \frac{1+|S_{11}|}{1-|S_{11}|} \tag{2}$$

At the output port, the VSWR (s_{out}) is given by

$$s_{out} = \frac{1+|S_{22}|}{1-|S_{22}|} \tag{3}$$

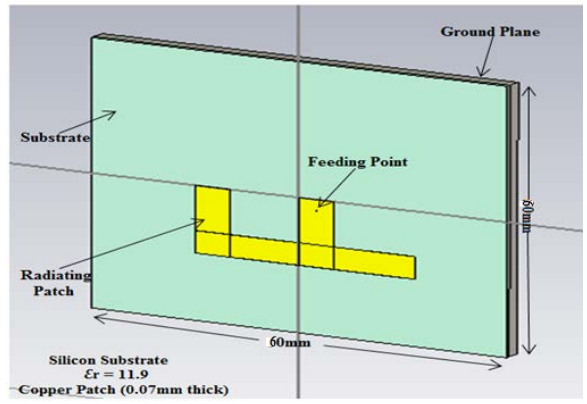
This is correct for reflection coefficients with a magnitude not greater than unity.

III. ANTENNA DESIGN

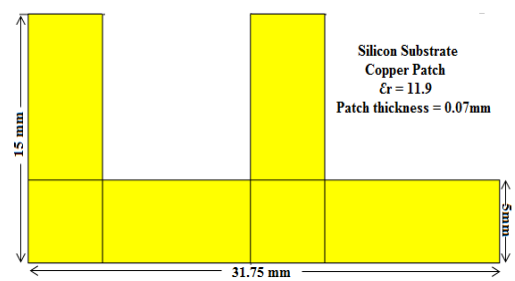
The basic PIFA consists of a ground plane, a top plate element, a feed of wire feeding, the resonating top plate, and a dc-shorting plate that is connecting the ground and the top plate at one end of the resonating patch. The IFA element is one of the most popular integrated antennas for plug-in interfaces due to its planar structure, small size and easy integration on a circuit board.

A. Structure of the Proposed Antenna

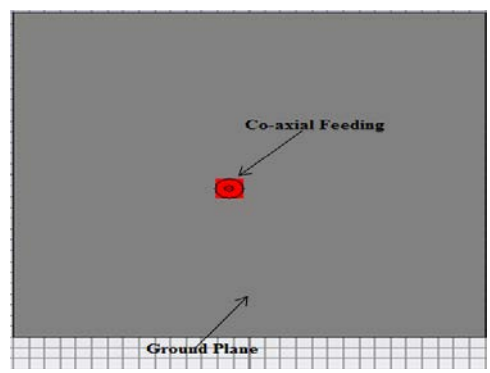
The proposed PIFA has a substrate is of the silicon material (ε = 11.9) and the substrate area is 60×60mm^2. Fig. 1(a) and 1(b) shows the structure of the PIFA element optimized for 2.45 GHz. Fig. 1(c) shows the co-axial feeding of the proposed antenna. In this structure the long arm of the patch is 31.75mm in length and the rest two arms are both 15mm in length and the width of each arm is 5mm. The height of the patch is 0.07mm. The ground plate is 2.1mm thick and the substrate plate is 0.7mm thick. Co-axial cable is used for the feeding of the antenna. The radius of the conductor is 0.5mm and the radius of outer insulator shield is 1.8mm. Waveguide port is set at the core of the co-axial cable in the direction of negative z-axis.



(a)



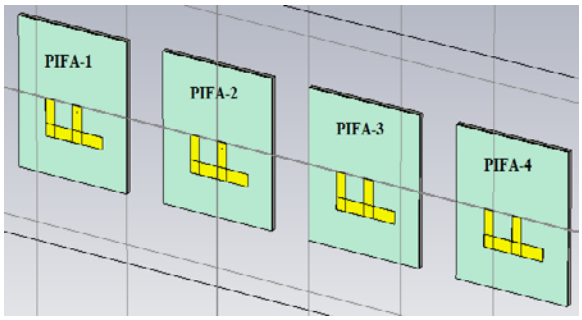
(b)



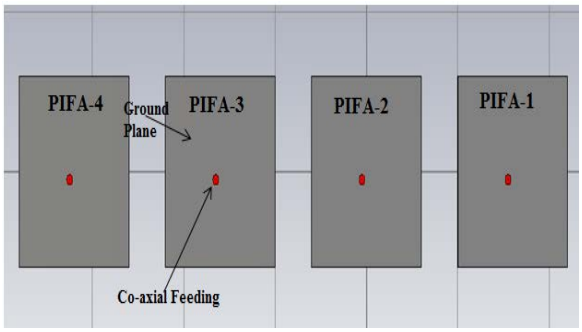
(c)

Fig. 1. Geometry of the single PIFA element (a) Perspective View. (b) Front View.

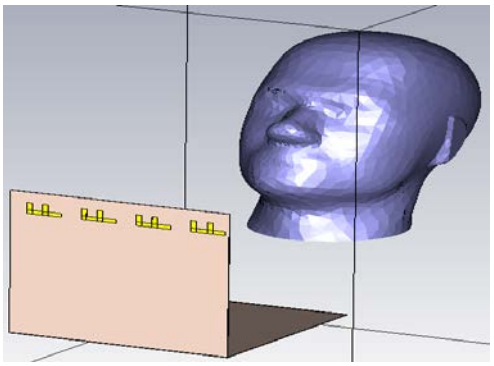
PIFA 4-element array is shown in Fig. 2(a). Fig. 2(b) and Fig. 2(c) show the co-axial feeding and antenna embedded on the laptop with phantom head model respectively. Four co-axial waveguide ports are used for the excitation of the antenna. These antennas can be excited simultaneously or individually.



(a)



(b)



(c)

Fig. 2. (a) PIFA array (perspective view) (b) Co-axial feeding line (Back View) and (c) Proposed Antenna embedded on Laptop with Phantom Head model.

IV. Simulation Results

A. S-parameter and Bandwidth

Fig. 3 focuses reasonable value of the S-parameter. The value at 2.45GHz is about -86.62dB. On the other hand, single element PIFA antenna has return loss value of about -67dB. From the figure, it can be seen that the antenna is an Ultra-wideband antenna. The antenna can be operated in a very wide range of frequencies ranging from below the 2GHz to above 3GHz scale.

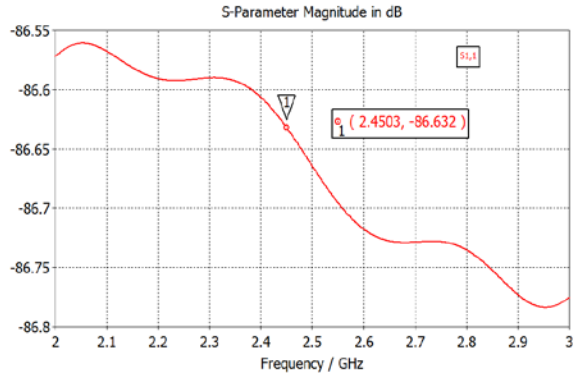
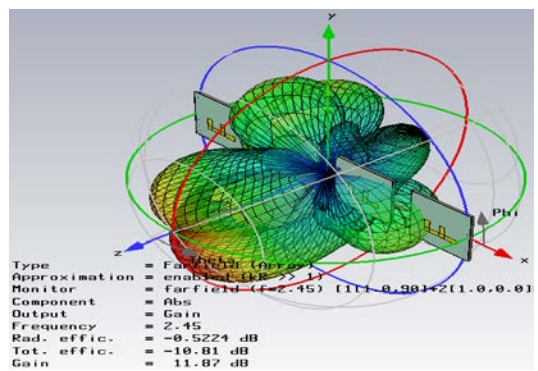


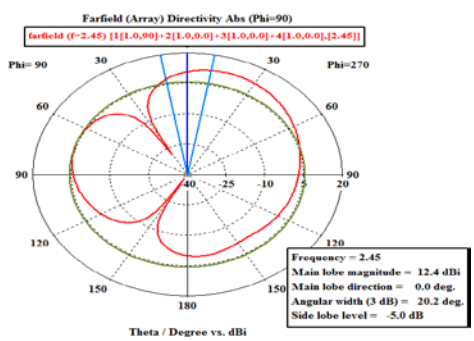
Fig. 3. S-parameter Magnitude versus Frequency.

B. Radiation pattern & Antenna Array Gain

The following Fig. 4 shows the far-field radiation pattern and gain of the PIFA 4-element array. From this figure it can also be seen that the gain of the antenna is about 11.87dB which is much better than the single element PIFA antenna. Single PIFA antenna gain is about 7.335dB.



(a)



(b)

Fig. 4. (a) Radiation pattern (b) Far-field Gain pattern.

The total antenna efficiency e_0 is used to take into account losses at the input terminals and within the structure of the antenna. Such losses may be due to the (i) reflections because

of the mismatch between the transmission line and the antenna, (ii) I^2R losses (conduction and dielectric). In general the overall antenna efficiency can be written as

$$e_0 = e_r e_{cd} = e_{cd}(1 - |\Gamma^2|) \qquad (6)$$

Where, e_r is the reflection efficiency $=(1 - |\Gamma^2|)$,

e_{cd} is the antenna radiation efficiency, which is used to relate the gain and directivity and the relation is given by

$$\text{Gain(dB)} = e_{cd}(\text{dB}) + \text{Directivity(dB)} \qquad (7)$$

$$e_{cd}(\text{dB}) = 10\log_{10}(e_{cd}) \qquad (8)$$

The antenna efficiency is calculated by these equations and the efficiency is 88.68%.

C. Voltage Standing Wave Ratio (VSWR)

The VSWR value of this array is very interesting. It is almost unity. Fig. 5 shows that the VSWR value is 1.0000932 at 2.45GHz.

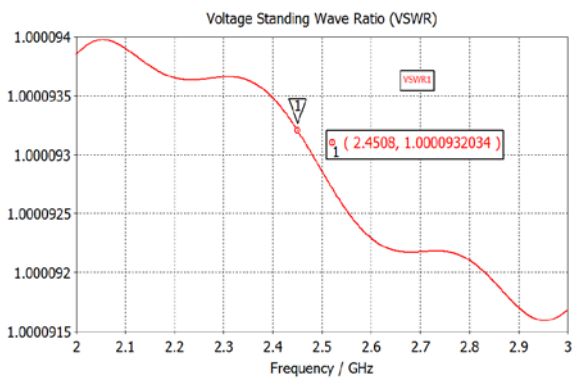
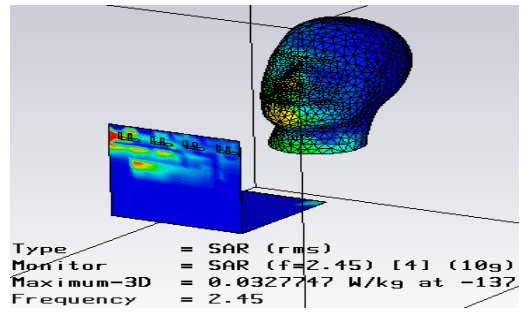


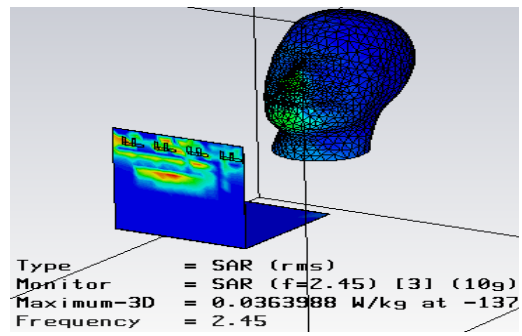
Fig. 5. VSWR versus Frequency (in GHz).

D. Specific Absorption Rate (SAR) and Antenna Efficiency

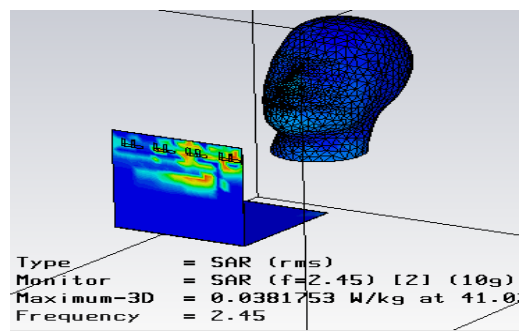
Fig. 6(a), 6(b), 6(c) and 6(d) shows the SAR values of the PIFA array for the excitation of 1st element (from left side), 2nd element, 3rd element and 4th element respectively averaged over 10g of tissue. Fig. 7 shows the SAR values of the PIFA array for the excitation of all 4-elements simultaneously. The FCC limit for SAR in the USA is 1.6 W/kg averaged over 1g of tissue. In Europe, the SAR limit is 2.0 W/kg averaged over 10g of tissue [1], [2]. From Fig. 6(a), 6(b), 6(c) and 6(d), the SAR values are shown 0.0327W/kg, 0.0363W/kg, 0.0381W/kg and 0.0684W/kg respectively averaged over 10g of tissue. In Fig. 7, the SAR value is shown 0.02075W/kg averaged over 10g of tissue.



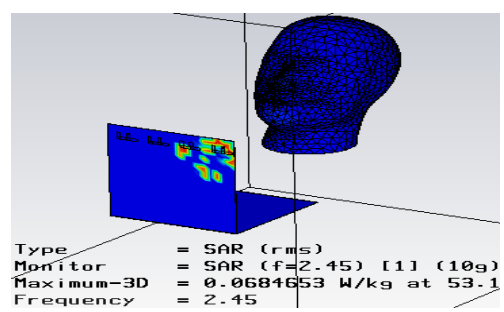
(a)



(b)



(c)



(d)

Fig. 6. SAR value for excitation of (a) 1st element (from left side). (b) 2nd element. (c) 3rd element. (d) 4th element (averaged over 10g of tissue).

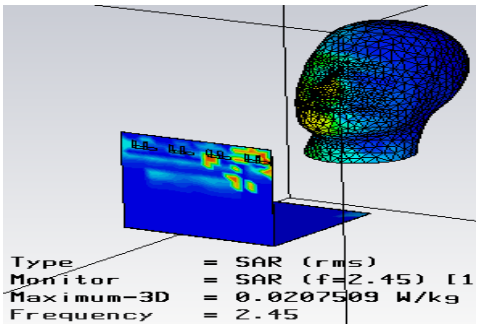


Fig. 7. SAR value for the excitation of all 4-elements simultaneously.

V. Result Analysis

The results of our proposed PIFA antenna array are obtained from the transient solver of the CST MW Studio 2011. The transient solver shows the values of antenna gain, return loss, VSWR, SAR and radiation pattern. The return loss value obtained by the solver is very interesting. The VSWR value is almost unity. The gain of the PIFA array shows a reasonable value. The SAR value for this PIFA array is very low which is far below the acceptance limit. The bandwidth is very wide. This wide band is very useful for high speed data transfer and communications. The efficiency of the antenna shows a very good value (88.68%).

TABLE I. Proposed Antenna Array Parameters and Their Values

Designed PIFA Antenna Parameters	Simulation Results
Voltage Standing Wave Ratio (VSWR)	1.0000927
S-parameter Value (in dB)	-86.62 dB
Gain (in dB)	11.87 dB
Directivity (in dBi)	11.92 dBi
3dB Angular Beamwidth (in deg.)	20.2 deg.
Antenna Efficiency (%)	88.68%

VI. Conclusion

The interaction of PIFA with the human head model has a significant effect on the radiation pattern, antenna gain and VSWR. However, the simulation result for the proposed PIFA antenna gives a very low value for peak SAR. The SAR value of the proposed antenna is less enough than the FCC limit. These values are better enough from other PIFA antennas. The antenna array has a very good figure of efficiency. This may be considered as an advantageous property of this PIFA antenna. This antenna can be used as a Wi-Fi antenna of the laptop which has a very wide bandwidth.

References

[1] ICNIRP (International Commission on Non-Ionizing Radiation), “Guidelines for limiting exposure to time-varying electric, magnetic, and electromagnetic fields (up to 300 GHz),” Health Physics, Vol. 74, 494–522, 1988.

[2] IEEE Std C95.1(tm)-2005, “IEEE standard for safety levels with respect to human exposure to radio frequency electromagnetic fields, 3 kHz to 300 GHz,” IEEE, New York, 2005.

[3] Guterman, J., Moreira, A. A., Peixeiro, C., and Samii, Y. R., “User interaction with Inverted-F Antennas integrated into laptop PCMCIA cards,” Radioengineering, Vol. 18, no. 2, June 2008.

[4] Guterman, J., Moreira, A. A., Peixeiro, C., and Samii, Y. R., “Antenna-user interaction in MIMO-enabled laptops,” Radioengineering, Vol. 18, no. 4, December 2009.

[5] Siaho, F. R. H., and Wong, K. L., “A shorted patch antenna with an L-shaped ground plate for internal mobile hand-set antenna,” Microwave and Optical Technology Letters, Vol. 33, No. 4, 2002.

[6] Guterman, J., Rahmat-Samii, Y., Moreira, A.A.,and Peixeiro, C., “Quasi-omnidirectional dual-band back-to-back E-shaped patch antenna for laptop applications”, IEEE Electronics Letters (Vol. 42 , Issue. 15), 28 August 2006.

[7] Antoniuk, J.,Moreira, A.A., and Peixeiro, C.“L-bent omnidirectional patch antenna for wireless applications in laptop computers”, IEEE Antennas and Propagation Society International Symposium, (Volume: 4A), 3-8 July 2005.

[8] Antoniuk, J., Moreira, A.A., and Peixeiro, C., “Multi-element patch antenna integration into laptops for multi-standard applications”, IEEE Antennas and Propagation Society International Symposium, (Volume: 4A), 3-8 July 2005.

[9] Chitra, R.J., Karthik, B.R., and Nagarajan, V., “Double L - slot microstrip patch antenna array for Wi-MAX and WLAN”, International Conference on Communications and Signal Processing (ICCSP), Chennai, India, 4-5 April 2012, pp. 137-140.

[10] Lin, S.-Y., “Very-wideband microstrip antenna for laptop applications”, Microwaves, Antennas & Propagation, IET (Vol. 5, Issue. 6), April 26 2011, pp. 705-710.

Monitoring the Performance of Computer User by Analyzing Physiological Signals

Md. Abu Baker Siddique Akhanda , Shaon Md. Foorkanul Islam and Md. Mostafizur Rahman
Department of Electronics & Communication Engineering
Khulna University of Engineering & Technology, Khulna-9203, Bangladesh
E-mail: Shumel0809@ymail.com, Shaon_ece_kuet@yahoo.com, mostafiz963@yahoo.com

***Abstract*—In this paper, we have analyzed the physical and mental performance of computer user due to the long term use of computer by analyzing the variations in physiological signals. Performances of computer user are monitored by recording and analyzing the variations in Electrocardiogram (ECG), Electromyogram (EMG), Electrooculogram (EOG), and Electroencephalogram (EEG). Detecting the performance of computer user throughout the computing session will enable more development of computer where it can interacts with its user taking users effective state in to account; known as affective computing. This research work has analyzed the variations in physiological parameters of a computer user which indicates the physical and mental condition of the user. All the signals are taken by using BIOPAC system and analyzed by using BIOPAC AcqKnowledge software and MATLAB. The different data sets of 12 subjects were collected where all the subjects were undergo a specific session of computer using which includes different computer mediated task. Research shows that there are significant variations in physiological signals throughout the session. This variation indicates differences in user physical and mental performance and can be used for affective computing purpose.**

Keywords— ECG; EMG; EOG; EE;, Mental state; Affective computing; BIOPAC; Acqknowledge software

I. INTRODUCTION

In this age of modern technology, people are greatly dependent on modern computer to reduce their work load and this electronic device, day by day has become the sole source of entertainment. People of all sort of age now a days use computer for different sort of application for a considerable amount of time. According to the Medical science Using computer for extended period of time are responsible for back problems, neck pain, arthritis, eye strain, bad posture, and any other negative effect from computer related stress [1] and here comes the application of affective computing. It is an advance field of engineering that attempts to provide computers with the ability to interact with its user and control its negative effect on the user [2]. So this situation has fostered to find out the variation in basic physiological parameters due to excessive use of computer and how they affect the performance of human body.

As the competition between industries are grows fiercer, human performance of task management are needed to be greatly improved. Human performances at working environment are already a great research topic [3]. Importance is mainly given to the stress performance. Computer mediated task performance calculation of computer user under stress and no stress condition are analyzed in physiological manner in [4]. Stress detection in computer users through non-invasive monitoring of physiological signals are done in it. Effect of mental loads on physiological signals is evaluated in [5]. Effect of color-word interference and task pacing on the physiological and emotional response are also studied [6]. But less attention has been paid to the computer user's physical and mental condition to the computer related task in physiological manner though most of the tasks now a day are performed by computer.

This performance measurement of the computer user can contribute greatly to the emerging discipline of Affective Computing. Computer can change the sound or colour according to the mental and physical condition of user adding new feature in human-computer interaction [7]. Until today main emphasis is given to the mental state classification by using biosignals [8]-[10]. Emotion recognition or classification of emotion also became a great research topic [11]. But less attention has been paid to verify mental performance in physiological manner during different computer mediated work condition. As we mentioned earlier goal of this research work is to verify this performance.

The novelty of this work is to make a decision how the physiological measurement varies with time and how they affect the physical and mental performance of human body when subjects are undergoing a specific session of computing. We have used ECG, EMG, EOG and EEG to find out these variations. A noteworthy problem is that physiological signals are affected by motion artifact. The fundamental application of this variability measurement is that, it will help us to decide how acutely our body and mind are affected by the excessive use of computer.

II. DATA ACQUISITION AND IT'S METHODOLOGY

A. Subjects

Physiological signals of 12 healthy subjects are collected by using BIOPAC software at Biomedical signal processing laboratory, Department of BME of EEE Faculty in Khulna

University of engineering and technology (KUET), Bangladesh. All the subjects are undergraduate student and are of good mental and physical condition.

B. Experimental process

The aim of this research work is to find out the effect of excessive use of computer and to find out the physical and mental performance of computer user by analyzing the variation of physiological signals. ECG, EMG, EOG and EEG were used as input measuring variable and variation of these parameters used to make decision about the performance of human body. For this purpose each subject was undergo a 240min(4h) of computer using session which includes internet browsing, typing, reading e-book, listening music and watching video. Physiological signals are collected from each subject in following manner.

TABLE I. TIME INTERVAL FOR VARIOUS SIGNAL COLLECTION IN 240MINS (4H)

Signals-> Time	ECG	EMG	EOG	EEG
1st hour	✓	✓	✓	✓
2nd hour	✓	✓	✓	
3rd hour	✓	✓	✓	
4th hour	✓	✓	✓	✓

ECG data were taken when the subjects are in sitting condition, using computer. EMG signals for the movement of hand muscle were taken when electrodes were connected in subject's right forearm and subjects were asked to type keyboard, browse or play computer games. EMG signal taken in this condition represents the strength of the skeletal muscle. EOG signals are taken during intense eye work like reading e-book, playing computer games. In this research work, subjects were asked to read e-book and signals were taken to evaluate the speed of eye movement from left to right (L-R) and right to left (R-L) direction to find out the fixation time of the eye which indicates the attention of the subject[12][13].

EEG signals were taken at beginning and end of the session to find out the mental condition of the computer user. For this purpose all the subjects were asked to perform some specific mental task.

- Resting state (RS): subjects in relaxed condition. EEG signals are taken for eye close and eye open condition.
- Mental test: All the Subjects were asked to solve some mental mathematical problem and given problems were increasingly tough order. Our goal was to observe the response of the subject to the problem and also to find the amount of time taken by the same subject to solve the same level of mathematical problem both at the beginning and end of the session. Importantly all this data were taken in eye close condition.
- Memory: A random number sequence like 5,66,7,16,4,9,2,0,3 was shown to all the subjects and subjects were asked to memorize the sequence. After giving 5mins subjects were asked to tell the number in sequential manner with eye closed and the corresponding signals were collected. This procedure was done to all subjects both at the beginning and end of session and goal was to compere two sets of data and make a decision about the memory condition.

Resting state with eye close condition data are taken as standard because Alpha wave amplitude of EEG is maximum with eye close condition and minimum with eye open condition. We use alpha wave to analyze because alpha wave amplitude varies with the amount of concentration given to task [9].

C. Data Acquisition

As we said earlier all the physiological variables are collected by using BIOPAC MP36 unit [1]. Required equipment's are BIOPAC electrode lead set (SS2L), BIOPAC disposable vinyl electrodes (EL503), BIOPAC data acquisition unit (MP36) with cable and power. Electrodes were placed on the subject's body according to the [1].

- ECG: White lead placed on right forearm, red lead on the left leg and the black lead was placed on right leg according to figure 1.
- EMG: All three leads placed on the right forearm shows in figure 3.
- EOG: Six electrodes are place on the front side of the face according to the figure 2.
- EEG: Red electrode was placed on the occipital lobe, the white electrode was placed behind ear and the ground black lead was placed on ear lobe. Practical EEG collection of a subject are shown figure 4.

Figure 1 & 2 are collected from [1]. Figure 3 & 4 shows the practical collection of muscle movement using EMG and EEG Alpha activity in BME laboratory.

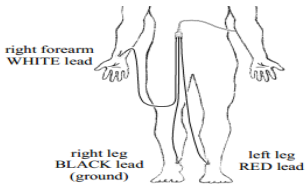


Fig. 1: Placement of lead for ECG

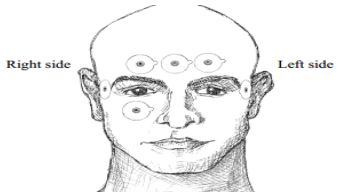


Fig. 2: Placement of lead for EOG

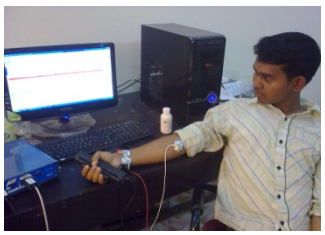


Fig. 3: Muscle movement measurement using EMG

Fig. 4: Alpha activity measurement using EEG

III. Data Analysis

As it is mentioned earlier all the recorded physiological signals are analyzed by using BIOPAC AcqKnowledge software and MATLAB. All the recorded physiological signals are converted in to frequency domain by Fast Fourier transform (FFT). Then the transformed data smoothed by removing noise. All the data are separately analyzed and analyzed data are compered to each other by using BIOPAC AcqKnowledge.

A. Fast Fourier Transform

A fast Fourier transform (FFT) is an algorithm that compute the Discrete Fourier transform (DFT) and its inverse. A Fourier transform is a process that converts time (or space) domain signal to frequency domain. FFT rapidly computes such transformations. As a result, fast Fourier transforms are widely used for many applications in engineering, science, and mathematics.

If $X(k)$ is one dimensional function of N elements, Then DFT is defined by the formula,

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-i2\pi k^{n}/_{N}} \quad Where\ k = 0 \ldots\ldots N-1 \qquad (1)$$

In this equation, k represents the harmonic number of frequency components. It transforms the time domain data into frequency domain. From Eq. (1), we get,

$$X(k) = \sum_{n=0}^{N-1} x(\mathrm{n}) \cos\left(2\pi \mathrm{n}\, {}^{k}/_{N}\right) - \mathrm{j} \sum_{n=0}^{N-1} x(\mathrm{n}) \sin\left(2\pi \mathrm{n}\, {}^{k}/_{N}\right) = R(k) - jI(k) \qquad (2)$$

Its magnitude can be calculated as,

$$|X(k)| = \sqrt{R^2(k) + I^2(k)} \qquad (3)$$

B. Smoothing

If "m" is the number of points in the window and "n" is the sample number The mean value smoothing formula is given by,

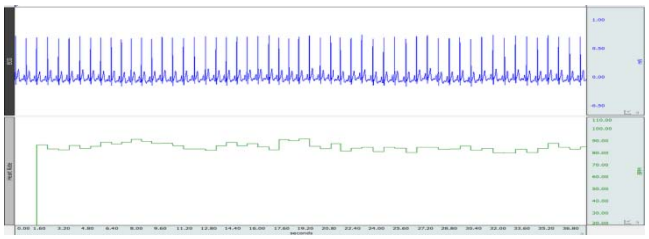
$$f(n) = \sum_{k=\,{}^{n-m}/_{2}}^{k={}^{n+(m-1)}/_{2}} f(k)\,/m \qquad (4)$$

AcqKnowledge software supports mean value smoothing and median value smoothing to remove noise around mean value and median value respectively. We use mean value smoothing with 100 sample frequency on every FFT transformed signal.

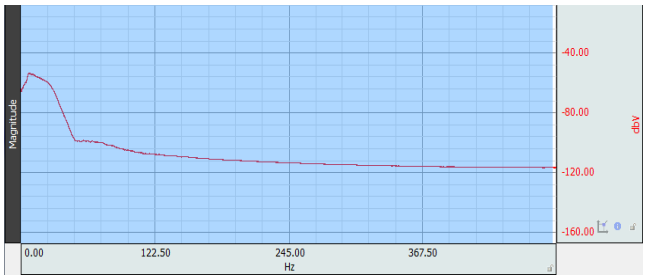
IV. Results

A. Recorded and transformed data

As we mentioned earlier in case of analyzing all the physiological signal resting state or data collected at the 1st hour was taken as the reference or standard data. Data taken in other stage are compared with 1st hour data for making decision. Figure 5-8 shows the FFT converted signals of each stage after smoothing. From the analysis of measured data it evident and described in the latter section that ECG signals don't give much variation with time in each stage. EMG, EOG, EEG signals shows important variation with each stage. So FFT figure of this three physiological signals of each stage after smoothing are given

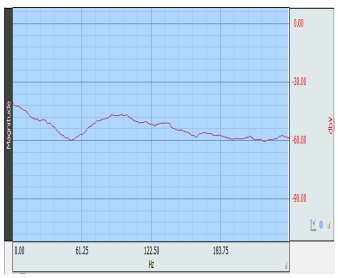


A: Recorded ECG and Heart signal

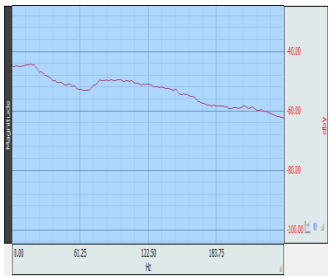


B: FFT data of ECG after smoothing

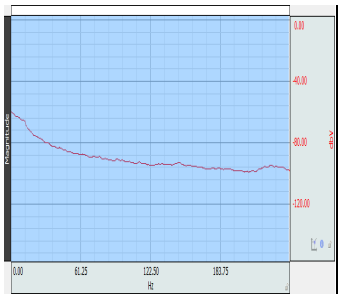
Fig. 5: ECG and Heart rate signal at 1st hour



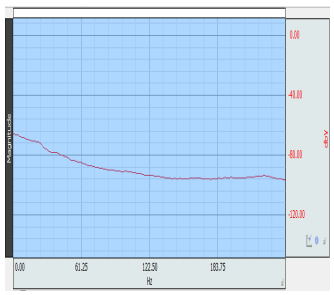
A: 1st hour FFT data of EMG



B: 2nd hour FFT data of EMG



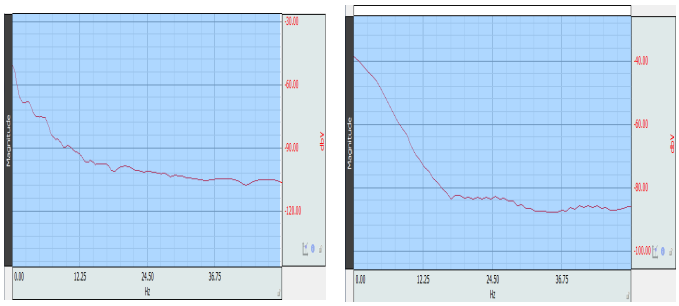
C: 3rd hour FFT data of EMG



D: 4th hour FFT data of EMG

Fig. 6: FFT signal of EMG after smoothing

In case of EMG, from the FFT data after smoothing, variation of Maximum amplitude, Minimum amplitude and Mean value in in dBV are taken as parameters to differentiate the changes in EMG signal at each stage of Signal recording. FFT signals are given in figure 6.

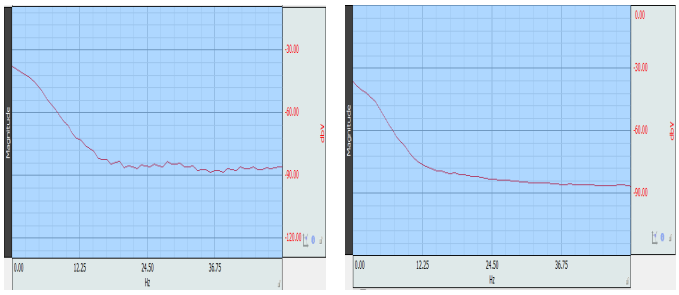


A: L-R movement of eye B: R-L movement of eye

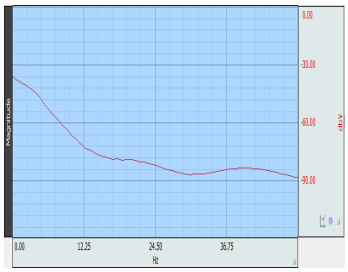
Fig. 7: FFT data of EOG signal for reading E-book at 1st hour

For analyzing EOG signal, Slope in 'dBV/hz' of the smoothed FFT signal and time taken in 'sec' for L-R and R-L motion of eye taken as parameter to calculate the variation of EOG signal at each stage. But both parameter are approximately remain constant for eye's L-R movement, so figure for 2nd , 3rd , 4th hour R-L motion of the eye are only given in figure 8.

For analyzing EEG data, Maximum amplitude in 'dBV' and frequency of max amplitude in 'Hz' of the smoothed FFT signal of Alpha wave are taken as parameter to find out the variation of data at each stage. Figure for RS,MT and MR of two data collecting stage are given in figure 9.

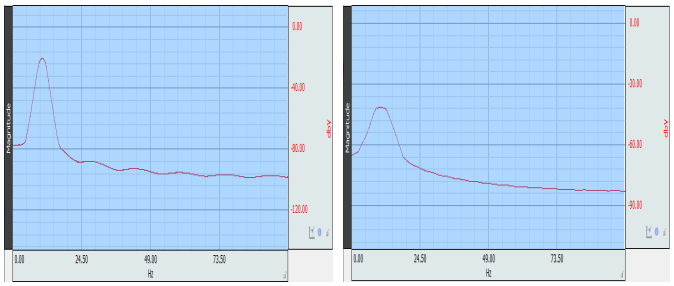


A: 2nd hour R-L FFT data of EOG B: 3rd hour R-L FFT data of EOG

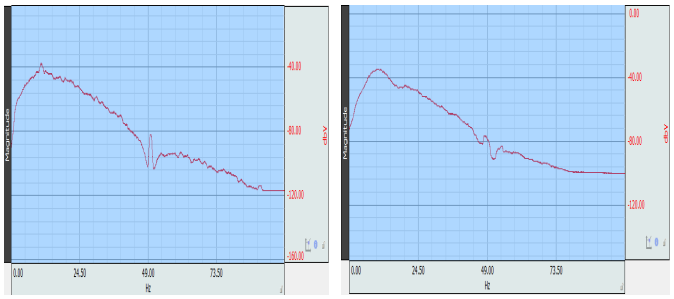


C: 4th hour R-L FFT data of EOG

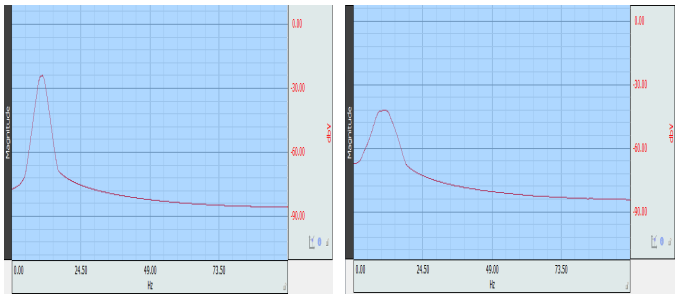
Fig. 8: FFT signal of EOG after smoothing



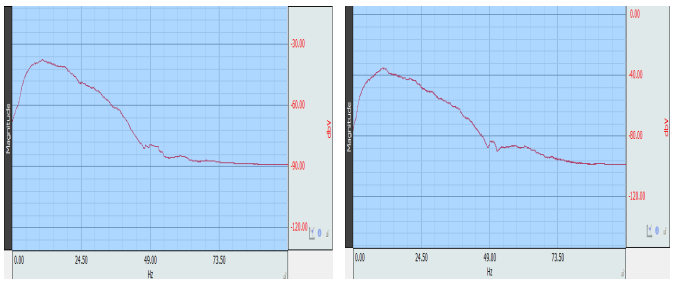
A: RS eye close state (1st hour) B: RS eye open state (1st hour)



C: Mental Test (1st hour) D: 1st hour Memory (1st hour)



E: RS eye close state (4th hour) F: RS eye open state (4th hour)



G: Mental test (4th hour) H: Memory (4th hour)

Fig. 9: FFT signals of EEG alpha wave in 1st and 4th hour after smoothing

B. *Results*

In this research, our goal was to find out the physical and mental performance of computer user by analyzing the variation in physiological signals. We have to collect a lot of data from each subject and analyze them thoroughly to evaluate the variation in physiological signal. In all the above smoothed FFT data, X-axis represents the frequency in 'Hz' and Y-axis represents amplitude value in 'dBV', where

dBV=$20log|X(k)|$. Now in case of ECG, data was analyzed for 8192 samples over 500 Hz frequencies in every stages. Experimental data are shown in TABLE II. From the data it is evident that there is not that much variation in ECG signal amplitude over entire session of data collection. Max dBV and max frequency are approximately constant. Heart rates are also in normal range.

TABLE II. ANALYZED RESULTS OF ECG SIGNAL AND HR

Time	Mean(BPM)	fmax(Hz)	dBVmax
1st Hour	82	8.259	-53.361
2nd Hour	71	7.90	-53.90
3rd hour	72	7.50	-54.28
4th Hour	69	7.40	-52.40

TABLE III. ANALYZED RESULTS OF EMG SIGNALS

Time	dBV max	dBV min	dBV mean
1st Hour	-40.42	-60.486	-50.38
2nd Hour	-44.38	-62.659	-55.36
3rd hour	-58.517	-86.402	-68.90
4th Hour	-65.542	-96.830	-68.50

In case of EMG, 1024 samples were analyzed over 250 Hz range in every stage and data for every 60 min are shown in TABLE III. From the data and Fig. 10 it is evident that max amplitude and power of EMG FFT data has great variation from 1st hour to 4th hour. And so does true to the mean value. As we mentioned earlier resting state or 1st hour data is taken as reference data. So the variations of EMG signal mean Amplitude and power increases in each stage and there is a strong loss of EMG signal strength as the increase in negative dB indicates less signal strength.

TABLE IV. ANALYZED RESULTS OF EOG SIGNALS

Time	Movement of eye	Slope(dBV/Hz)	Time needed (sec)
1st hour	L-R	-1.1400	6.02
	R-L	-0.96276	0.65
2nd hour	L-R	-1.1600	6.21
	R-L	-0.9805	0.81
3rd hour	L-R	-1.2000	6.58
	R-L	-1.01963	1.02
4th hour	L-R	-1.1900	6.75
	R-L	-1.0600	1.22

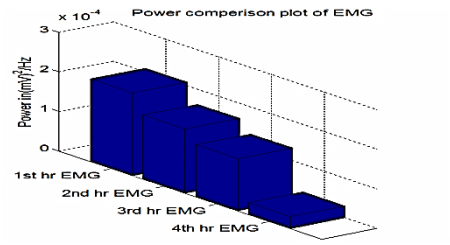


Fig. 10: Power comparison of 4hour EMG signals

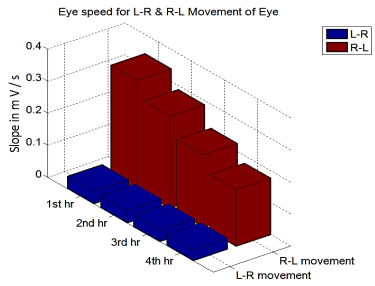


Fig. 11: Eye speed for L-R and R-L movement of eye

EOG analyzed data were given and plot in to the TABLE IV and Fig. 11 where *slope* in 'dBV/Hz', represents speed of eye movement from Left to Right and Right to Left direction. '*Time needed'* in 'sec' for L-R, represents the time taken by a subject to read a single line from left to right side and R-L represent vice versa. From the data and Fig. 11 it is evident that eye's speeds to move from L-R remain approximately same but R-L varies increasingly from 1st to 4th hour. And time needed to complete a single line and move from one stop to next line (fixation time) increases at each stage.

TABLE V. ANALYZED RESULTS OF EEG ALPHA SIGNALS

Time	RS eye close		Memory (dBV max)		Mental test (dBV max)		RS eye open (dBVmax)	
	dBV max	F max (Hz)	dBV max	Fmax (Hz)	dBV max	F max (Hz)	dBV max	F max (Hz)
1st hr	-21.70	10.5	-41.10	9.9	-21.70	10.5	-41.10	9.9
4th hr	-23.51	10.7	-37.01	10.5	-23.51	10.7	-37.01	10.5

TABLE VI. MENTAL TEST AND MEMORY STATISTICS

Data stage	Mental test (Avg. time needed to complete test)	No. of subjects able to complete test	No. of subjects not able to complete	MR (Avg no. of "numbers" memorized by subjects)
1st hour	115sec	10	2	5
4th hour	133sec	9	3	4

In case of analyzing EEG signal, as we mentioned earlier unlike other physiological signal in this research EEG signals were taken at the start and end of the total session. Data's are collected from each subject for Resting State (RS), Mental

Test (MT) and Memory (MR). Calculated data's are shown in TABLE V and total power of alpha wave is plotted in Figure 12.

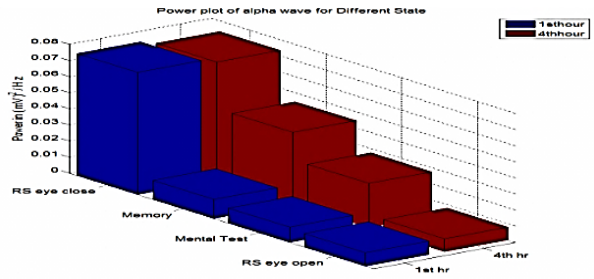


Fig. 12: Power comparison of Alpha wave for different mental state throughout the data collection stage

In this research we only analyze the amplitude of alpha wave of EEG signal to find out the variations at different state data in two stages of data collection. From the TABLE V and Fig. 12 Alpha wave amplitude and total power is maximum when the subject in RS eye close and minimum for RS eye open because Alpha activity is minimum when eyes open or attending to stimuli [9]. Alpha activity is inversely proportional to the concentration of subject to the task [9]. From data and Fig. 11 Alpha activity of MT and MR in 1st hour are much lower than 4th hour Alpha activity approximately equal to the value of RS eye open state, indicates much more concentration in 1st hour. TABLE VI shows average time taken by the subjects to complete Mental Test in 1st hour & 4th hour and number of subjects able to complete the test in two stage of data collection. It also shows the average number of "numbers" sequentially memorized by 12 subjects. Now the above data are taken for the smoothing factor 100. But this data can be varied according to the variation of smoothing factor [9].

V. Conclusion

In this research work, we tried to find out the variations in both mental and physical performance of a computer user by the variation of physiological signals for long term computing. Results show a less variation in ECG signal but a significant variation in signal strength for EMG and EOG signal. For EOG data, R-L movement speed of eye decreases in each hour and corresponding fixation time increases. Time lengths of fixations indicate attention of the subject to the task perform (i.e, reading). For EEG, alpha activity is max when subject was in eye close RS and minimum for eye open state both in 1st and 4th hour. But analysis of variation in alpha wave in two stages shows that in case of MT and MR, 1st hour data has lower amplitude and power than 4th hour. As Alpha activity is inversely proportional to the concentration, this result indicates subjects have more concentration in 1st hour than 4th hour. Again average time needed to complete the test was higher and number of person able to complete the test was lower in 4th hour (133s & 9 subject) than 1st hour (115s and 10 subject).For Memory, Average "5 numbers" are memorized sequentially by 12 subjects in 1st hour and in 4th hour it is "4 numbers". These results also indicate that subjects have less concentration and efficiency in MT and MR in 4th hour than 1st hour which describes the variation of mental performance of subject throughout computing session. However it was tough to collect data accurately because of the environmental condition and motion artifact. But we give our full effort to the fidelity of our data collection. In future we will involve other physical variable like Galvanic Skin Response (GSR), Blood Volume Pulse (BVP) and Skin Temperature (ST) and will try to analyze the signal using neural network for better result.

References

[1] Reference manuals of AcqKnowledge Software Guide and MP System Hardware Guide, BIOPAC Systems Inc., 42 Aero Camino, Goleta, CA – 0067.

[2] R. W. Picard, Affective Computing. Cambridge,MA: M.I.T. Press, 1997.

[3] D.J.Abramis, "Work role ambiguity, job satisfaction, and job performance: Meta-analysis and review". Psychol. Rep. 75, 1411–1433 (1994).

[4] N. Yong Jo , K. Chang Lee and D. Sung Lee, "Computer-Mediated Task Performance Under Stress and Non-stress Conditions: Emphasis on Physiological Approaches" K.C. Lee (ed.), Digital Creativity: Individuals, Groups, and Organizations,Integrated Series in Information Systems 32, DOI 10.1007/978-1-4614-5749-7_2, © Springer Science+Business Media New York 2013.

[5] K. Iwanga, S. Saito, Y. Shimomura, H. Harada, and T. Katsuura, "The Effect of Mental Loads on Muscle Tension, Blood Pressure and Blink Rate", Journal of Physiological Anthropology and Applied Human Science, vol. 19, no. 3, pp. 135 – 141, 2000.

[6] P. Renaud and J.P. Blondin, "The stress of Stroop performance: physiological and emotional responsesto color-word interference, task pacing, and pacing speed," International Journal of Psychophysiology, vol. 27, pp. 87-97, 1997.

[7] E. Hudlicka, "To feel or not to feel: The role of affect in human-computer interaction," International Journal of Human-Computer Studies, vol. 59, pp. 1-32, 2003.

[8] M. Takahashi, O. Kubo, M. Kitamura, and H. Yoshikawa, "Neural Network for Human Cognitive State Estimation", Proceedings of the IEEE Conference on Intelligent Robots and Systems, Advanced Robotic Systems and Real World, IROS, vol. 3, pp. 2176 – 2183, Sept 1994.

[9] T. T. K. Munia, A. Islam, M. M. Islam, S. S. Mostafa and M. Ahmad "Mental States Estimation with the Variation of Physiological Signals" ,IEEE/OSA/IAPR International Conference on Informatics, Electronics & Vision (ICIEV) 2012.

[10] C. Zong, M. Chetouani, "Hilbert-Huang Transform Based Physiological Signals Analysis For Emotion Recognition", IEEE International Symposium on Signal Processing and Information Technology ISSPIT, 2009.

[11] C. K. Lee, S. K. Yoo, Y. J. Park, N. Kim, K. Jeong, and B. C. Lee, "Using Neural Network to Recognize Human Emotions from Heart Rate Variability and Skin Resistance", Proc. of the 27th IEEE Annual Conference on Engineering in Medicine and Biology, Shanghai, China, pp. 5523-5525, September 1-4, 2005.

[12] K. Rayner, J.Pollatsek and B.Alexander, "Eye Movements During Reading." The science of reading: A handbook.[1-4051-1488-6]: Blackwell Publishing. pp. 79-97. ISBN 1-4051-1488-6 (Hardcover); 978-1-4051-1488-2 (2005).

[13] J. Wang, "Pupil Dilation and Eye-tracking." A handbook of process tracing methods for decision research: a critical review and user's guide: Society for Judgment and Decision Making Series. pp. 185-204.ISBN1-84872-864-6(2011).

A Novel Audio Watermarking Algorithm Based On Neighboring Samples

Shahid Md. Asif Iqbal[1], Alok Kumar Chowdhury[1], Amina Akhter[2], Kingshuk Dhar[1]
[1]Department of Computer Science & Engineering, Premier University, Chittagong, Bangladesh
{asifcsep, alokchy04, kingshuk2006}@yahoo.com
[2]Department of Computer Science, Asian University for Women, Chittagong, Bangladesh
runeycse99@yahoo.com

Abstract

A simple and effective audio watermarking algorithm based in time domain is proposed in this paper for controlling the unauthorized distribution of copyright audio. The algorithm makes use of the advantages of both embedding a watermark bit to an individual sample and embedding a bit to a group or frame of samples. Here, a frame of three positive valued samples is used to define the amount of distortion to be added but only the middle sample is distorted based on the decision. Based on the watermark bit either "0" or "1", the middle sample of a frame is made equal to one of the neighboring samples. Thus, the watermarked audio has a very good perceptual quality. Experimental results show that our new audio watermarking algorithm is not only imperceptible, but also robust against various common signal processing attacks such as amplitude modification, re-sampling, low-pass filtering, re-quantization, MP3 compression, etc. Though proposed in time domain but the algorithm can be implemented in any transform domains with little or no modification.

Keywords: Distortion, Imperceptible, Frame, Neighboring samples, Middle sample.

I. INTRODUCTION

As soon as the world's music industries diverted from the analog music to digitally encoded samples the threats of unauthorized reproduction and illegal distribution of the original works are frustrating content owners or proprietors of the music. Popularity, low-price, usefulness and easy availability of Internet based multimedia systems and widespread deployments of computers are adding more security threats to the copyright protection. Thus, to prevent the digital audio from unauthorized copying digital audio watermarking was introduced and through these years it is performing well in protecting the copyright information by injecting the owner's signature or symbol in the distributed audio signal. Digital watermarking is an information embedding technique that inserts imperceptible and virtually undetectable information into digital data (e.g. video, images and audio signals). Information embedding should be done such that embedded information (any significant characters, texts, or images) is hidden and it bears the unique presence of the owner or the distributor. Digital data authors and distributors rely on digital watermarking to claim the ownership of their scholarly property rights at the same time permit other individuals to legally distribute digital data. The embedding must be robust to common signal processing and compression techniques and various intentional attacks.

According to the International Federation of Phonographic Industry (IFPI), audio watermarking should meet the following requirements:

•**Imperceptibility:** The digital watermarked audio signal should have a good quality .i.e. the amount of distortion made to the original audio signal should be insignificant.

•**Robustness:** The embedded watermark data should be conserved or extractable despite of various asynchronous attacks (amplitude modification, re-sampling, re-quantization, MP3 compression, etc.) and common audio signal processing operations or compressions.

•**Capacity:** Capacity refers to the number of bits (watermark information) that can be embedded into the audio signal within a unit of time.

•**Security:** Security implies that only the authorized person can detect or extract the watermark data by statistical analysis.

Among the above requirements imperceptibility and robustness are important than others. These two requirements are often contradictory with each other and a trade-off between them is needed for an imperceptible and robust watermarking. These contradictory objective requirements pose many challenges to achieve a good quality audio watermarking.

Satisfying the two important properties of good audio watermarking we propose a simple and computationally less complex watermark embedding and extracting algorithm. Our algorithm mainly uses three neighboring samples to form a frame and then embed watermark data only in the middle simple. The embedding causes the middle simple to be equal to either the first or last sample thus adjusting or distorting the sample by a little amount which is undetectable and invariant against most of the attacks.

II. RELATED WORKS

Watermarking techniques [1-15] proposed during the last ten to twelve years can be divided into two broad categories, namely the time-domain based group ([9] [12] [15]) and the transform-domain based group. The transform-domain based works ([1]-[8], [10]-[11], [13] – [14]) mainly use cepstrum domain transform, discrete Fourier or frequency transform (DFT), discrete cosine transform (DCT), discrete wavelet transform (DWT), etc.

The technique proposed by Lee et al. in [1] inserts watermarks into the cepstral components to hide a narrow band signal in a wideband channel. Multiple scaling parameters are used to adjust the audio signal whereas fixing the threshold value for finding the first scaling parameter was left open. Zhang et al. [2] have deployed cepstrum domain transform to use block encoding that performs well when high recovery rate is required but suffering lower communication rate. The performance degrades with the increase in the number of bits transformed.

Quantization index modulation (QIM), distortion-compensated QIM (DC-QIM), and dither modulation proposed by Chen et al. [3] seem to achieve better imperceptibility–robustness trade-offs than spread-spectrum and low-bit(s) modulation methods. Lihua et al. [4] proposed a watermarking algorithm based on DWT which makes use of the quantization length D for better outcomes but selecting the best value of D using the host signal or other applicable parameters is a major challenge. Mean quantization of cepstrum coefficients are used by the authors in [5] which modifies the minor cepstrum coefficients near the centre and suffers from the same problem as above.

Using the self-error-correction capability of BCH code and attack-defensive nature of cepstrum domain the algorithm proposed by the authors in [6] seems to exhibit robustness against most of the asynchronous attacks. However, the BCH code itself can correct only a finite number of errors. BCH code based composite technique [7] combining cepstrum transform-discrete wavelet transform shows lower bit error rate (BER) of the extracted watermark.

In the paper [9], Lie et al. proposed an embedding method in time domain that makes use of differential average-of-absolute-amplitude relations within each group of samples (GOS) to represent one-bit information. The algorithm maintains high security of the embedded watermarks employing two keys, however fails to detect the cropping attack made at the middle or last part of the signal.

Modified audio signal keying (MASK) [12] tried to modify the short-time envelope of the audio signal using characteristics curves to select optimal set of parameters for specific applications. Thus, the performance of the scheme is depended on the intelligent selection of the set of parameters.

Ren et al. [13] used a 24 bit colour image as the watermark to enhance capacity. Using two-dimensional Arnold transform, then DCT, and wavelet transform the watermarking information was embedded into the wavelet low frequency components. The algorithm is imperceptible and robust.

Authors in [15] proposed a faster technique based on Schur decomposition and Dither Modulation whose performance is depended on the image size and user-driven positive real numbers e.g. quantization number and dither value.

The QIM based techniques [3], [4], [5], and [11] all suffer from intelligent determination of the quantization steps. Other embedding algorithms also suffer from defining the amount of distortion when watermarks are embedded. But, our algorithm determines the distortion amount very easily and simply while keeping the watermarked signal perceptually inaudible.

Transform-domain based techniques are complicated and not trivial to implement. The time domain based algorithms are conceptually simple and easy to implement. So, we propose an easy to implement, robust and imperceptible algorithm mainly in the time domain but which can also be applied in the other transform do-mains with little or no modification.

III. PROPOSED ALGORITHM

The time-domain based watermarking algorithms embed watermark information by adjusting each sample, with a disturbing quantity, of the host audio signal independently [9]. However, distorting an audio sample independently without considering the neighboring samples may lead to a watermarked signal with audible noise or perceptible modification if the amount of distortion is high. On the other hand, grouping or framing consecutive samples to embed watermark information scales the amplitudes of the samples of the group in a consistent manner. Though, it leads to imperceptible watermarked audio signals but determining the size of the group or frame is not trivial and the technique may handle the asynchronous attacks poorly. The former techniques increase the watermark embedding capacities hiding one watermark bit in one sample, but suffer from high noisiness if the distortion amount is not adjusted properly. The latter embed less watermark bits by embedding one bit either “1” or “0” to a group of samples so that failing to detect watermark of an individual sample does not necessarily cause a wrong decision concerning watermark retrieval. But, the techniques may suffer from pirate attacks if the distance between “0” and “1” states is not sufficient. Thus, the difference between the two states is important to maintain the optimum audio quality [9]. Hence, we make a trade-off by adjusting or modifying only a single sample but taking the distortion amount from the two neighboring (behind and front) samples

A. Watermark Embedding Algorithm

We describe the basic steps involved in the proposed watermark embedding algorithm below

Step I. The host audio signal, sampled at 44100 Hz, is divided into non-overlapping frames, with a frame size of three (3) samples. The three samples should be positive and the first and last samples should be unequal. Thus, the distance between consecutive frames is not same.

Step II. A binary image of size **M × M** given by **P = {p(m1, m2): 1≤m1≤M, 1≤m2 ≤M, p(m1,m2)∈ {0, 1}}** is used to generate the watermark. We convert the two-dimensional binary image into one-dimensional sequence of **M × M** bits in order to create the watermark to embed into the audio signal. Thus, the watermark is W_{em} **= {w(n)∈ {0, 1}: 1 ≤ n ≤ M × M}**.

Step III. The $\mathbf{j}^{th}$ sample of the $\mathbf{i}^{th}$ frame is given by $\mathbf{x_{i,j}}$. Thus $\mathbf{x_{i,1}} \neq \mathbf{x_{i,3}}$ and **i ≤ n**. The watermark bit is embedded in the middle sample of each frame.

Step IV. If a frame is found satisfying the constraints mentioned in step 1, then depending on the value of watermark bit **w(n)** as found from W_{em} following procedure is used to embed watermark in the middle sample of the frame:

$$\textbf{if}\,(w(n) == 0)$$
$$x_{i,2} = x_{i,1}, \quad if\,(x_{i,1} \le x_{i,3})$$
$$x_{i,2} = x_{i,3}, \quad otherwise$$
$$\textbf{else}$$
$$x_{i,2} = x_{i,1}, \quad if\,(x_{i,1} \ge x_{i,3})$$
$$x_{i,2} = x_{i,3}, \quad otherwise$$

B. Watermark Extraction Algorithm

The extraction of the embedded watermark can be performed without the host signal. The basic steps involved in the watermark extraction process are given below:

Step I. The watermarked audio signal is segmented into non-overlapping frames, with a frame size of three (3) samples. The three samples should be positive and the first and last samples should be unequal.

For each frame **i** calculate $\mathbf{x_a}$ using eq. (1)

$$x_a = (x_{i,1} + x_{i,3})/2 \tag{1}$$

The distortion amount **D** is thus given by eq. (2)

$$D = |x_{i,2} - x_a| \tag{2}$$

The bit value of the extracted watermark W_{ex} is determined by the following procedure

$$\textbf{if}\,(x_{i,2}) \le x_a$$
$$e(n) = 0$$
$$\textbf{else}$$
$$e(n) = 1$$

Thus, W_{ex} **= {e(n)∈ {0, 1}: 1 ≤ n ≤ M × M}**

IV. Experimental Results & Analysis

We tested our algorithm on different types of audio signals which were 16 bits signed mono sampled at 44.1 kHz in WAVE format. We have used different types of audio signals like Pop, Country, and Classical music. A plot of a small portion of the pop audio signal and the corresponding watermarked classical audio signal is plotted in Fig. 1. In all the experiments conducted, we have used a 36 × 20 binary image as our watermark, which is displayed in Fig. 2.

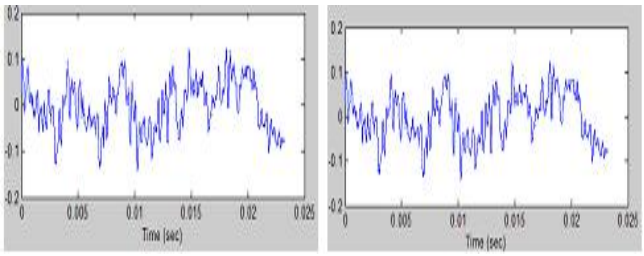


Fig. 1. Portion of original audio and watermarked audio signals

Fig. 2. Watermark binary image

To determine the imperceptibility or quality of our watermarked audio signals we have used signal to noise ratio (SNR) and is given by the eq. (3).

$$SNR = 10\log_{10}\frac{\sum_{n=1}^{N} X_o^{\,2}(n)}{\sum_{n=1}^{N}[X_o(n) - X_w(n)]^2} \tag{3}$$

where X_o and X_w are the host audio signal and watermarked audio signal respectively. After embedding the watermarks, the SNR of all used audio signals are above 48 dB. Thus, we can say that our embedded watermark is perceptually inaudible. Our algorithm shows imperceptibility because the amount of distortion makes the middle sample of a frame equal to one of the neighboring samples. Since, the natural pattern of an audio signal is that the consecutive or adjacent samples have little difference in their amplitudes; our embedding algorithm preserves the audio envelope and achieves good quality audio.

To validate the robustness of our algorithm, we performed the following five different types of attacks to the watermarked audio signals:

- **Amplitude Modification:** Various types of amplitude modification attacks such as 3dB boost, 3dB cut, 5% and 10% increase in amplitude are applied here.
- **Re-quantization:** The 16-bit watermarked audio signals have been re-quantized down to 8

bits/sample and up to 32 bits/sample and finally back to 16 bits/sample again.

- **Re-sampling:** The watermarked signal originally sampled at 44.1 kHz is re-sampled at 32.0 kHz and also at 96.0 kHz, and finally restored by sampling again at 44.1 kHz.
- **Low-pass filtering:** The low-pass filter with cut-off frequency of 8 kHz and 12.5 kHz are applied on the watermarked audio signal.
- **MP3 Compression:** The MPEG-1 layer 3 compressions with 128 kbps, 256 kbps and 320 kbps are applied on the wav format of watermarked audio signal and back to wav format again.

The normalized correlation (NC) is used to evaluate the similarity measurement of extracted binary watermark, which can be defined using eq. (4):

$$NC(W_{em}, W_{ex}) = \frac{\sum_{p=1}^{M}\sum_{q=1}^{M} W_{em}(p,q) W_{ex}(p,q)}{\sqrt{\sum_{p=1}^{M}\sum_{q=1}^{M} W_{em}^{2}(p,q)}\sqrt{\sum_{p=1}^{M}\sum_{q=1}^{M} W_{ex}^{2}(p,q)}} \quad (4)$$

where W_{em} and W_{ex} are embedded and extracted watermarks respectively and p, q are indexes of binary watermark image.

The BER is used to find the percentage of errors made during the extraction of the watermark i.e. the error bits between the embedded watermark and extracted watermark. The BER is given by eq. (5):

$$BER = \frac{B}{M \times M} \times 100\% \quad (5)$$

where B is the number of bits incorrectly detected.

Table 1 shows the extracted watermark with various NC values. The extracted watermark image is almost similar to the original watermark image.

Table 1. The extracted watermark with various NC values

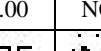
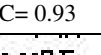
NC= 1.00	NC= 0.93	NC= 0.91
		
NC= 0.96	NC= 0.87	NC= 0.99

Table 2 gives the performance of our algorithm. The low BER and high NC between extracted watermark and original watermark indicate that our proposed scheme has very high robustness.

Robustness performance of the proposed algorithm is mainly depended on the distortion amount **D**. The distortion in the samples of the watermarked audio signals caused by disturbances or attacks is recognized by **D** i.e. our algorithm can detect the right watermark in presence of the distortion amount of **D**. Our scheme can recover exactly same watermark image which was embedded when amplitude modification attacks were applied and almost in all case the NC value is 1.00. Though re-quantized down to 8bits/sample and again back to 16bits/sample losses significant amount of information our scheme can recover over 75% of the watermark image. The NC value against other attacks was above 0.80, which indicates the high performance of our scheme. We can also observe that the audio files like Classical, Country, and Pop stand out as good host signals for our watermarking scheme, because under above attacks watermark recovery is not disturbed significantly.

Table 2. NC, BER and SNR of our algorithm against various attacks

Song	Attack	Parameter	NC	BER	SNR
Pop	No Attack	No Attack	1	0	48.6991
	Amplitude	-3dB cut	1	0	
		+3dB boost	1	0	
		+5%	1	0	
		+10%	1	0	
	Re-quantization	16-8-16	0.842857	25.9722	
		16-32-16	1	0	
	Re-sampling	44.1-32-44.1	0.895867	17.6389	
		44.1-96-44.1	0.931601	11.6667	
	Low Pass Filter	8 KHz	0.709892	44.0278	
		12.5KHz	0.776858	35.2778	
	MP3	128kbps	0.861408	23.0556	
		256kbps	0.875573	20.6944	
		320kbps	0.896162	17.5	
Classical	No Attack	No Attack	1	0	61.3501
	Amplitude	-3dB cut	1	0	
		+3dB boost	1	0	
		+5%	1	0	
		+10%	1	0	
	Re-quantization	16-8-16	0.755798	39.1667	
		16-32-16	0.891234	18.6111	
	Re-sampling	44.1-32-44.1	0.907984	16.1111	
		44.1-96-44.1	0.887627	19.1667	
	Low Pass Filter	8 KHz	0.813013	29.1667	
		12.5KHz	0.849419	24.8611	
	MP3	128kbps	0.895263	18.3333	
		256kbps	0.904624	16.6667	
		320kbps	0.898221	17.6389	
Country	No Attack	No Attack	1	0	59.7607
	Amplitude	-3dB cut	0.917953	14.0278	
		+3dB boost	0.999188	0.138889	
		+5%	1	1	
		+10%	1	1	
	Re-quantization	16-8-16	0.777363	35.2778	
		16-32-16	0.87896	20.6944	
	Re-sampling	44.1-32-44.1	0.936716	10.9722	

		44.1-96-44.1	0.966593	5.69444	
	Low Pass Filter	8 KHz	0.849553	27.0833	
		12.5KHz	0.875257	20.6944	
	MP3	128kbps	0.914106	14.8611	
		256kbps	0.896104	17.7778	

V. Conclusion & Future Research Directions

A new simple and efficient method of audio watermarking algorithm has been proposed in this work. Here, we tried to make a trade-off between embedding one bit in each individual sample and embedding to a group of samples. We used three consecutive positive samples to make a frame and then based on the watermark bit either "0" or "1" the middle sample of the frame is made equal to one of the two neighboring samples. Since, the watermarked sample is changed to one of its neighboring sample the imperceptibility of our algorithm is very high. The experimental results show that the proposed method has good capabilities of robustness and inaudibility. In addition, computation and implementation our proposed scheme is very straight forward method. However, there are scopes for further improvement. Here, we have considered only the samples which are positive. But, we are almost at the end of extending the work by taking more samples, both positive and negative, in a frame. In that extension, we are also considering the critical situation of the first and last samples of a frame being equal.

REFERENCES

1. S. K. Lee, Y. S. Ho, "Digital Audio Watermarking in the Cepstrum Domain," IEEE Transactions on Consumer Electronics, Volume 46, Issue 3, pp. 744 – 750, AUGUST 2000.
2. L. ZHANG, Z. HUANG, X. CAO, "A Digital Audio Watermarking Algorithm Based On Block Encoding," International Conference on Wireless Communications and Signal Processing, pp. 1 – 5, November 2011.
3. B. Chen, G. W. Wornell, "Quantization Index Modulation: A Class of Provably Good Methods for Digital Watermarking and Information Embedding," IEEE Transection on Information Theory, Volume 47, Issue 4, pp. 1423 – 1443, MAY 2001.
4. M. Lihua, Y. Shuangyuan, J. Qingshan, "A New Algorithm for Digital Audio Watermarking Based on DWT," WRI Global Congress on Intelligent Systems, Volume 4, pp. 229 – 233, May 2009.
5. B.K. Vivekananda, I. Sengupta, A. Das, "Audio Watermarking Based on Mean Quantization in Cepstrum Domain," 16th International Conference on Advanced Computing and Communications, pp. 73 – 77, December 2008.
6. S. C. Liu, S. D. Lin, "BCH Code-Based Robust Audio Watermarking in the Cepstrum Domain," Journal of Information Science and Engineering, Volume 22, Number 3, pp. 535 – 543,May 2006.
7. B.K. Vivekananda, I. Sengupta, A. Das, "Audio Watermarking Based on BCH Coding using CT and DWT," International Conference on Information Technology, pp. 49 – 50, December 2008.
8. I. K. Yeo, H. J. Kim, "Modified Patchwork Algorithm: A Novel Audio Watermarking Scheme," IEEE Transactions on Speech and Audio Processing, Volume 11, Issue 4, pp. 381 – 386, July 2003.
9. W. N. Lie, L. C. Chang, "Robust and High-Quality Time-Domain Audio Watermarking Based on Low-Frequency Amplitude Modification," IEEE Transactions on Multimedia, Volume 8, Issue 1, pp. 46 – 59, Feb. 2006.
10. Q. Guo, Y. Zhao, P. Cheng, F. Wang, "An Audio Digital Watermarking Algorithm Against A/D And D/A Conversions Based on DCT domain," 2nd International Conference on Consumer Electronics, Communications and Networks, pp. 871 – 876, April 2012.
11. W. Tan, S. Yang, Y. Chen, J. Zhou, "Research on DFT Domain Digital Audio Watermarking Algorithm Based on Quantization," First International Workshop on Education Technology and Computer Science, Volume 3, pp. 736 – 739, March 2009.
12. A. N. Lemma, J. Aprea, W. Oomen, L. van de Kerkhof, "A Temporal Domain Audio Watermarking Technique," IEEE Transection on Signal Processing, Volume 51, Issue 4, pp. 1088 – 1097, April 2003.
13. K. Ren, H. Li, "Large Capacity Digital Audio Watermarking Algorithm Based on DWT and DCT," International Conference on Mechatronic Science, Electric Engineering and Computer, pp. 1765 – 1768, August 2011.
14. X. Li, H. H. Yu, "Transparent and robust audio data hiding in subband domain," International Conference on Information Technology: Coding and Computing, pp. 74 – 79, March 2000.
15. A. Choudhary, S. P. S. Chauhan, S. Tanveer, M. A. Alam, "Schur decomposition and dither modulation: an efficient and robust audio watermarking technique," Proceedings of the CUBE International Information Technology Conference, pp. 744 – 748, September 2012.

Mobile Phone SIM Card: A Security Concern in the Perspective of Bangladesh

Nazmul Hasan[1], Mohammad Riasat Ahmed[2], Raisul Islam[3], Nazia Majadi[4]
[123]Department of Computer Science and Engineering
[123]Military Institute of Science and Technology, Dhaka, Bangladesh
[4]Lecturer, Department of Computer Science and Telecommunication Engineering
[4]Noakhali Science and Technology University, Noakhali, Bangladesh
E-mail: [1]nazmul_mist@yahoo.com, [2]neon_rpc@yahoo.com, [3]raisul_54_7182@yahoo.com, [4]nazia_majadi@yahoo.com

Abstract— **At present time the unregistered SIM cards (here unregistered means the SIM cards which are registered with false information provided by dealer) are widely used for committing crimes in the perspective of Bangladesh. Due to the complexity of the present SIM card registration process, the subscribers are reluctant to register themselves. In this paper we propose a cloud based system model which demonstrates an online method of SIM card purchasing and registration. This system model conquers the constraints of the conventional system of mobile purchasing and registration which is lengthy and stalking. It enables the subscriber to buy the SIM card and to be registered in an uncomplicated way. Here we mainly emphasize on the security issue through mandatory mobile SIM card registration. The purpose of our system model is lessening the rate of crimes which can be taken place by the use of mobile. We also show that how this system will encourage the people to be a registered user and at the same time discourage the criminals to commit crime with the use of mobile call and SMS.**

Keywords- **Authentication, cloud computing, central database, PIN code, SIM.**

I. INTRODUCTION

Mobile phones have become part of our everyday life. Now-a-days, mobile phones have vastly changed the way we communicate with each other. A mobile phone can be all we need for communicating. From a mobile phone we make calls, send text messages, emails, send and receive directions, go on the internet, buy things, do online banking, listen to music and much more. Bangladesh is a small country with six different mobile operators (Grameen phone, Robi, Teletalk, Airtel, Citycell and Banglalink) and a huge number of subscribers get service from them. Each and every day the number of subscribers is increased. At the end of May, 2013 total number of mobile subscribers of Bangladesh is 102.995 million which is 17 lac 90 thousand more than the number of users of April, 2013 (101.225 million) [2]. Along with Bangladesh, only 12 countries in the world have more than 100m active subscribers. BTRC claimed that the mobile penetration of Bangladesh is 66.36% among the whole population with a growth of 10% per annum.

Cell phones have positive effect in mitigating crimes as it allows for quicker reporting of crimes, and, in some cases, real time communication of details about the crime and the criminal that helps the victim to escape. On the other hand cell phones have also a negative impact on crime [3]. Various numbers of crimes involve the use of cell phones (Calls & SMSs) such as- collecting toll, mugging, demanding extortion after kidnapping, planning and executing of robbery and stealing activities, blackmailing, issuing threat to kill or other harm, fomenting militancy etc [4]. These crimes rate are augmented alarmingly with the rising figures of mobile subscribers. Most of these crimes are occurred with the use of unregistered SIM cards as it helps criminals to hide their real identity because the information against that SIM card can be inserted with false information with the help of the dealer and make the task very difficult for the law enforcing agencies to detect their whereabouts [8]. So these unregistered or fake registered SIM cards can be used as weapon by the criminals to perform their illegal actions. According to a report, up to October, 2009, there were 10 million SIM cards used for crime in Bangladesh [7].

The previous caretaker government (2007-08) took steps for the SIM card registration and framed rules for this purpose. But an unspecified number of people at that time used fake documents for registration of SIM cards. In March, 2010 Bangladesh officials have blocked 1.5 million mobile numbers and tightened regulations for buying and selling SIM cards following a spike in mobile-related crime. In these blocked 1.5 million numbers; 200,000 phones were being used in cases of extortion and

978-1-4799-3497-3/13/$31.00 ©2013 IEEE

hasslement while 1.3 million numbers were blocked because they lacked full registration documents [5]. Usually each criminal or terrorist keeps 400 to 500 unregistered SIM cards at his disposal. A SIM card is destroyed after using it only once. Though BTRC has a consumer protection guideline [6], but those are difficult to implement without knowing the actual information about the users. Our government doesn't get the tax for the unregistered SIM as it is not documented and the operator can easily hide the transaction's information.

Operators	Active Subscribers (in millions)
Grameen Phone Limited (GP)	43.263
Banglalink Digital Communication Limited	26.574
Robi Axiata Limited (Robi)	22.192
Airtel Bangladesh Limited (Airtel)	7.662
Pacific Bangladesh Telecom Limited (Citycell)	1.408
Teletalk Bangladesh Limited (Teletalk)	1.897
Total	**102.995**

Figure 1.1**:** Mobile phone subscribers at the end of May, 2013 [1].

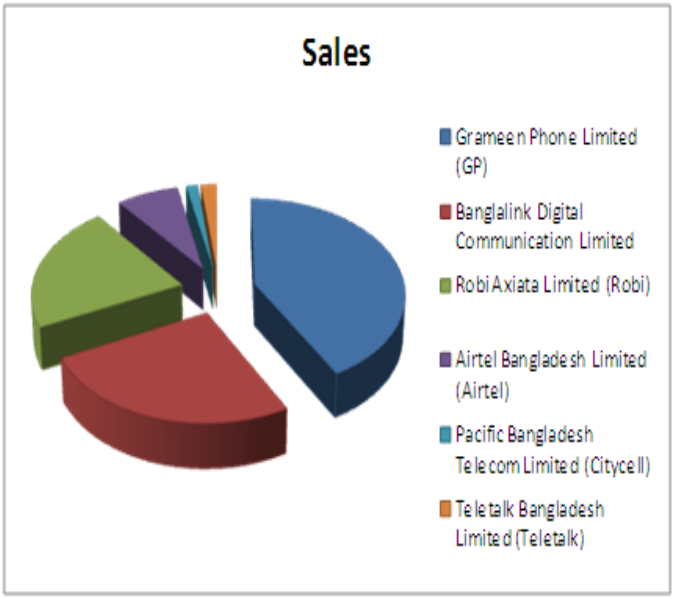


Figure 1.2: Pie chart of the mobile phone subscriber.

The rest of the paper has been organized as follows: Section II provides related research, System Model is given in Section III, Section IV shows the performance analysis and Section V describes recommendation. Finally conclusion is provided in Section VI.

II. RELATED RESEARCH

There is a very limited number of research works based on this topic. But a good number of newspaper article, magazine article and scientific blog article have found to relate this topic. Among them a blog named natunbarta.com have analysis on the number of mobile SIM card subscriber in Bangladesh. Another web blog *eHOW*™ have released an article about impact of mobile phone on crime. Online newspaper site *bdnews24.com* has published news on the activities of BTRC against crime using cell phone. Another online newspaper the hidustantimes has made a detail report on the crime by using mobile phones in Bangladesh.

III. SYSTEM MODEL

As the crime occurs by using SIM card cannot be traced due to the wrong information given at the purchasing time. At present the manual SIM card registration process is active. In this system, it is very easy to get a SIM card by providing wrong information. So we have proposed the centralized cloud based SIM card purchasing / registration system. There will be a central database containing information for all the citizens of the country. The profile of each citizen will create during birth registration time, so each citizen will have a profile with a unique ID number (like the national ID card number for the citizen in Bangladesh) in that database. For this system, we will need three types off user, they are the customer (who wants to purchase the SIM card), the dealer of the respective telecommunication operator (who is authorized to sell the SIM card), and a body of authority (who will be responsible for adding and deletion of the mobile number in the individuals profile).

Figure 2: Three types of users in the system.

A. *Architecture for the proposed system*

There will be central database containing the information of all the citizens, dealer and the authority. The personal information is to be inserted at the birth registration. In the individual's profile, there will be the name of the person, address, national ID number, and the mobile phone numbers. Among these information, the profile owner only can update the password of his profile, rest of all the information will be maintained and update centrally. There will be another option as "Generate Code" for set a security code and its validity period. This security code is very important as it will be used later for accessing this profile by the SIM card dealer. There is a message option for getting the SIM card activation PIN code along with the mobile phone number.

For the dealer's profile, there is an option to communicate with user, like searching for a customer profile for verification by using the customer ID of that customer along with the security code which was set by the customer himself. There is another option for the dealer to send the PIN code of the SIM activation to the customer profile. There is an option for dealer to communicate with authority. The dealer can request for deletion for a number from any person's profile upon the user of that mobile number surrender the SIM card to the dealer. The dealer ID is set by the respective mobile phone operator.

For the authority, s/he can update his/her password for the own profile. There is a message option for this user to receive a request from the dealer for deletion a mobile phone number from any user's profile.

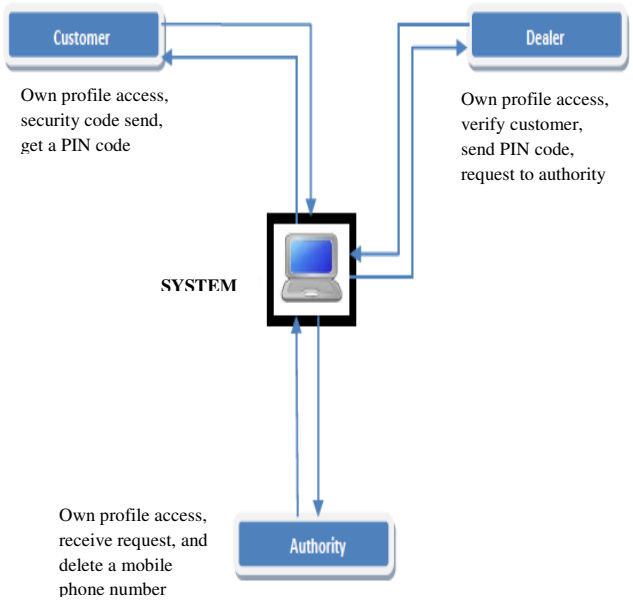


Figure 3: Context diagram of the system

B. *Purchasing a SIM card*

When a person wants to purchase SIM card, at first he has to log in into his profile and set security code with validation period. Then he will go to the desired mobile operator's sell center. He will give his national ID number and the security code which was set by him to the dealer. The dealer will verify the customer with customers' national ID number and security code. Then the dealer will give the SIM card to the customer, but the SIM card is in inactive mode. The dealer will send a SIM activation PIN number to the customer's profile along with the mobile number. Then the person will log in into his profile and see the message to retrieve the mobile number and the PIN code. By using this PIN code, the inactive SIM card become active, and at the same time the number will automatically add to the user's profile.

Thus the total number of the SIM cards will be noted in his profile. So the subscriber is being enrolled by the mobile phone operator at the same time government has the knowledge which subscriber has purchased which SIM card from which dealer.

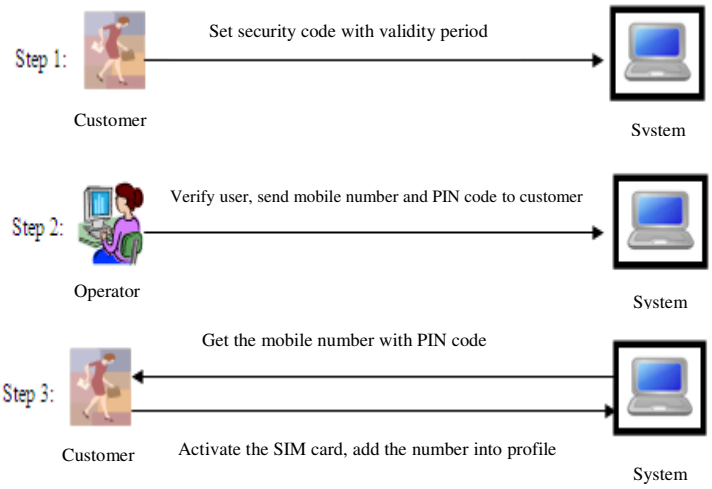


Figure 4: Steps of purchasing and activating SIM card

C. *Surrender a SIM card by the user*

A user can surrender his SIM card to the operator due to some specific reason. For this operation, the user set a security code to his profile, and goes to the dealer with a written statement stated that, why he surrendering the SIM card. The dealer will verify this user by the previous procedure. Then the dealer will send a message to the authority with the surrendered mobile number and user

national ID. The authority then deletes the mobile number from the user's profile.

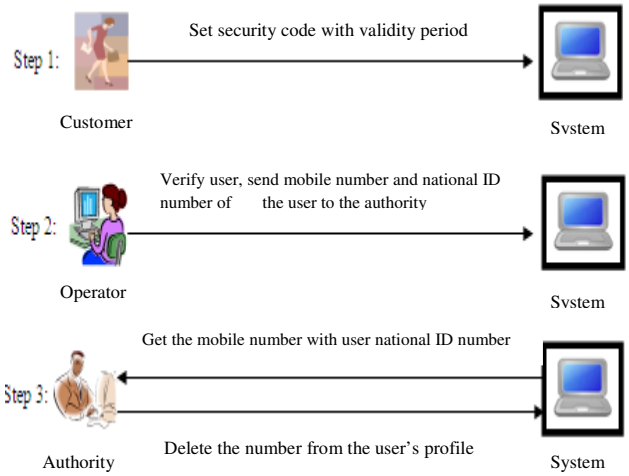


Figure 5: Steps of surrender a SIM card.

The results of this system are proper documentation of the SIM card subscription, identification of the SIM card subscriber and dealer, accurate statistics of the cell phone users, number of SIM card used by a single person, uncomplicated but secured information transaction during SIM card purchasing. Reduce the time consumption and avoid the uses of the hard copy of any document, proper authentication of the customer is assured. It dispels the deterrence of having SIM card at under age (below 18) as is introduces central database system instead of national ID card for SIM card purchasing.

The output of this system is very significant as it is concerned about the security regarding the involvement of mobile phones in criminal activities. It will reduce the rate of the crimes by using the cell phones because, every SIM card will be recognized by its subscriber, and will be identified very easily. If the dealer of any mobile phone operator involves with criminal activities by providing wrong information to help any subscriber, s/he can be easily identified by this system. The government can easily identify the statistics of the cell phone users. As the number of sold SIM card is monitored, so the government can earn proper tax from this sector. As the total number of SIM card used by an individual can be recognized and generally criminals are used to use a number of SIM cards at their disposal, so security control authority can suspect and trace anyone by the number of their SIM cards. Besides infamous criminals can be traced as they cannot hide their contact information from their profile.

IV. PERFORMANCE ANALYSIS

The rate of cyber crime using mobile phone has become too acute that, Directorate General of Forces Intelligence (DGFI) Bangladesh considers it an alarming sign. They are worried that fake registration of mobile phone SIM card, especially by criminals will increase in near future amid current political crisis. It has stated earlier that, BTRC have blocked 1.5 million SIMs which is about 2% of the total subscriber for fake registration upon complaints from Rapid Action Battalion (RAB) and Police. [9] So it can easily be assumed that, there would be some proportion (assume X %) of the rest 98%, despite have not lodged any complaint yet have gone though fake registration process. But by the use of this proposed model these X% will be deceased.

On fake registration of SIMs, Dhaka Metropolitan Police's Deputy Commissioner Masudur Rahman recently told "When we contact operators about fake registration they say retailers have not yet sent back the registration forms, so they cannot offer any details."[9] As the proposed system offers real time online registration process compulsory to purchase a SIM card, so this kind of excuse from operator will not be greeted.

Mobile operators usually claim that they had no access to the national database which is under the Election Commission – and they have no alternative tools to detect real identification of the subscribers. In this proposed system, the operators have to access the central database that contained the details information of the subscribers. Besides the existing database contains information of the people older than 18 years, but the proposed central database will be enriched with the information of all the people of Bangladesh as it allows getting a SIM without of age restriction, because it may required to have a mobile phone to the people younger than 18 years.

A case is being considered by tracing a mobile number that created any crime by the International Mobile Station Equipment Identity (IMEI). But TIM Nurul Kabir, secretary general of Association of Mobile Telecom Operators of Bangladesh said "the IMEI barring is a tool to ensure security but SIM registration should be real before that."[9] Because for a same SIM card several handset can be used. According to this, the system stated in this paper, ensure SIM card registration process in such a way that the user of the SIM card can be recognized without difficulty and dispel the necessity of being bothered the IMEI number.

In the existing SIM card transaction process, the dealers are too concern to assure their performance bonus assign on selling on a specific amount of SIM card. For this reason, they often provide SIM card to the customer without taking any kind of information from them to attain their satisfaction and involved themselves in fake registration process. But by the proposed model such kind of the activities are not possible, as the dealers information is also stated during the registration process.

V. RECOMMENDATIONS

As this proposed system will work under a central database, so the management of that database is very crucial. The accurate information must enter after through checkup. The data entry authority must be trustworthy and efficient. The security of the database is another important concern. There should be enough security for the database. The proper backup must be ensured.

At present we have made a demo software to represent the system's functions. The system has been implemented in a short range. The experimentation of the whole system is going on very sharply. If the cordial support is confirmed by the respective authority, this system can be tested and deployed in the practical field.

VI. CONCLUSION

In this paper we have proposed a cloud based SIM card registration process which can be implemented by the use of a central database. Our proposed system is hassle free for both the customer and the dealer. Our target is to enlist all the subscribers through this online registration process that reflects the exact number of subscribers. At the same time it will ensure the proper documentation of necessary information needed to identify each and every subscriber. So, 100% recuperation of taxes by the tax authority of government will be ensured. Any criminal can be traced by the SIM card used in criminal activities as no one can be able to purchase SIM card without giving the correct information. We are optimist that this method mitigates the crime rate very effectively. The system is organized by gathering some cryptographic process to ensure the authentication. That is why our proposed system is better from the present manual SIM card registration process. Although we have proposed the model for Bangladesh but it can be implemented to any other country.

ACKNOWLEDGMENT

The authors are highly grateful to Gp.Capt. Md. Afzal Hossain, head of the department, Department of Computer Science and Engineering, Military Institute of Science and Technology, Dhaka, Bangladesh, and Md. Kishwar Shafin, undergraduate student, Department of Computer Science and Engineering, Military Institute of Science and Technology, Dhaka, Bangladesh.

REFERENCES

[1] Mobile Phone Subscribers statistic, *telekothon.com*, http://www.telekothon.com/2013/06/mobile-phone-subscribers-at-end-of-may-2013.html [last visited: 08/07/2013]

[2] 103mn mobile subscriber in Bangladesh, *natunbarta.com*, http://www.natunbarta.com/english/si-tech/2013/07/01/5948/103mn-mobile-subscribers-in-Bangladesh [last visited: 10/07/2013]

[3] The Impact of Cell Phone on Crime, *eHOW™*, http://www.ehow.com/about_5398414_impact-cell-phones-crime.html [last visited: 09/07/2013]

[4] BTRC moves to combat crime using cell phone, *The Financial Express*, http://www.thefinancialexpress-bd.com/index.php?ref=MjBfMDlfMzBfMTJfMV84OF8xNDUyODM= [last visited: 12/07/2013]

[5] Bangladesh blocks 1.5 million mobile phones to fight crime, *news.com.au*, http://www.news.com.au/technology/bangladesh-blocks-1-5-million-mobile-phones-to-fight-crime/story-e6frfro0-1225839922680 [last visited: 08/07/2013]

[6] BTRC moves to protect cell phone users, *bdnews24.com*, http://bdnews24.com/business/2013/05/24/btrc-moves-to-protect-cellphone-users [last visited: 12/07/2013]

[7] 10 mn SIM cards used for crime, terror in Bangladesh: Police, *The Hindustan Times*, http://www.hidustantimes.com/world-news/Bangladesh/10-mn-SIM-cards-used-for-crime-terror-in-Bangladesh-Police/Article1-461103.aspx [last visited: 09/07/2013]

[8] Illegal SIM cards being misused by terrorists? *CiOL*, http://www.ciol.com/ciol/news/87125/illegal-sim-cards-misused-terrorists [last visited: 11/07/2013]

[9] DGFI concerned about crime through mobile phone networks, DhakaTribune, http://www.dhakatribune.com/bangladesh/2013/nov/02/dgfi-concerned-about-crime-through-mobile-phone-networks [last visited: 07/11/2013]

Adoption of E-Banking in Bangladesh: Evolution, Status and Prospects

Salekul Islam, ASM Mahfuz
United International University, Dhaka, Bangladesh
Email: salekul@cse.uiu.ac.bd, asm.mahfuz@gmail.com

***Abstract*—Electronic banking (e-banking) is the next-generation banking that electronically delivers a wide range of banking services where the physical presence of the customer at the bank premises is not required. With the advent of Internet connectivity and a number of security tools of e-commerce, the penetration of e-banking is increasing most of the developing countries, e.g., Bangladesh. Although there are different forms of e-banking, including Internet banking, telephone banking, mobile phone or Short Message Service (SMS) banking and Automatic Teller Machine (ATM) banking; Internet banking is the most cost-effective form of e-banking. The number of Internet users is geometrically increasing in Bangladesh. Moreover, the banks have started to invest in e-banking, specially in Internet banking. Although, Internet banking has immense potential in Bangladesh, it fails to attract the customers in general. In this paper, by carrying out an on-line survey, we study the existing e-banking facilities of Bangladesh, the use of this e-banking services and the challenges that need to be addressed for full-scale deployment of e-banking. By analyzing survey data we have identified the key reasons behind the slow spread out of e-banking services, and also prepared a number of recommendations to overcome the present hindrances.**

Index terms–E-banking, e-commerce, customer satisfaction, survey data analysis, e-banking adoption.

I. Introduction

Electronic banking (e-banking) is the way of delivering any banking services through electronically [1]. It encompasses a wide range of banking services including all the services–traditional brick-and-mortar bank–can provide. Depending on the delivery channels [2], e-banking can be categorized in different groups, such as Internet banking, telephone banking, mobile phone or Short Message Service (SMS) banking, Automatic Teller Machine (ATM) banking, etc. The rapid growth of the usage of the Internet and the declining cost of the Internet bandwidth have emerged the Internet banking [3] the most popular and promising way of e-banking. Although, the Internet banking is a subset of e-banking [1], among the general users, there is a common misconception of considering it the only way of e-banking and thus, using it interchangeably with e-banking.

Although, e-banking has been around for many years through ATMs and telephone banking, more recently it gets the momentum due to the wide spread use of the Internet. Both customers and financial institutions could be benefitted through e-banking. Customers can access fast and convenient services around the clock, even from a remote place. On the other hand, banks and financial institutions can provide services more economically and efficiently. For example, a typical customer transaction costs about $1 in a traditional brick-and-mortar bank branch, while the same transaction costs $0.60 through a phone call and only about $0.02 through Internet banking [4]. Therefore, there exists vast demand of e-banking, specially Internet banking from both parties of the stakeholders (i.e., the customers and the banks).

E-banking has been successfully commenced in the developed countries many years ago: the first ATM machine started to operate in the USA in 1959 and the first home online banking services was set up in 1983 in the UK. However, in the developing countries the spread of e-banking is much limited due to many socioeconomic and technological factors including: lower penetration of information and telecommunication technology, the lower computer literacy, the 'digital divide' between the rich and poor, the different operational environments for public and private sector banks, problems of security and authentication, unfriendly law and regulation [5]. Note that not all these factors are equally manifested in every developing country and thus, for a specific region/country, the obstacles in the growth of e-banking need to be identified first. In connection to this fact, the three research objectives of this paper are:

1) Summarize all the e-banking services provided by different banks in Bangladesh and also identify which services are being frequently used by the customers.
2) Analyze the usage pattern and identify the hindrances responsible for under utilization of some services.
3) Finally, provide some recommendations for the rapid growth of the overall e-banking services in Bangladesh.

To determine which e-banking services are being used frequently and also why customers prefer those services over other available services, we carry out a comprehensive, online survey. More than 80 participants, both male and female from a diverse age groups and educational background take part in the survey. By analyzing the survey data, we successfully identify the frequently used e-banking services and also determine why some other services failed to attract customers. The evaluations of the existing e-banking services will guide the administrators of the banks to re-engineer their services and also assist them to decide which services should be given more attention in future.

II. Related work

The development, adoption and prospects of e-banking in Bangladesh has been investigated by a number of researchers.

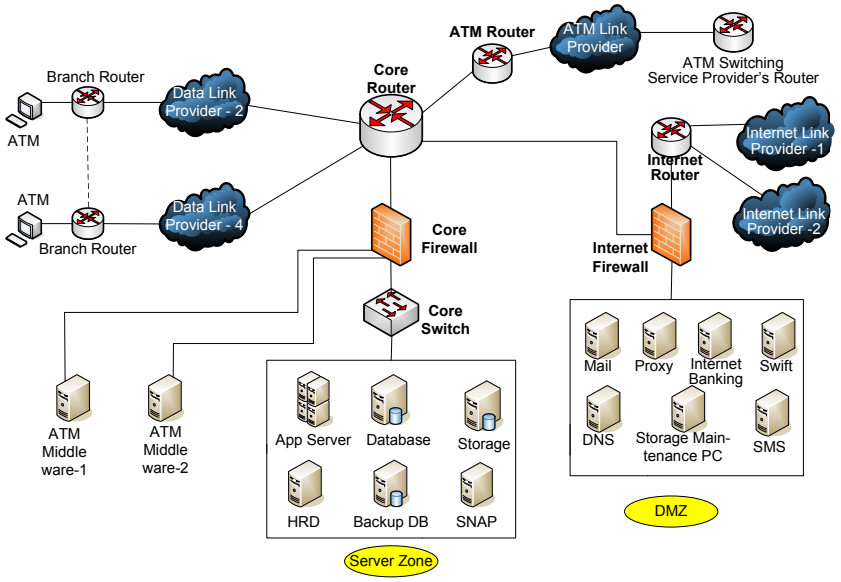


Fig. 1. The architecture of the banking network of Bangladesh

Here, we present the summary of some studies. Since, most of the developing countries have similar situations, in some studies, the scenario of Bangladesh has been considered as a case study. In [6], the researchers have found that the lack of infrastructure is the major issues for Internet banking. An empirical study to investigate the recent trend and development of the application of e-banking in Bangladesh has been presented in [7]. In the marketing point of view a survey-based study has been presented in [8]. The study demonstrates that the Bangladeshi customers have not enough knowledge in e-banking and the benefits they may receive through it. Small and medium enterprise (SMEs) might be benefitted through e-banking by reducing transaction costs and establishing greater control over bank accounts. A study to determine the relationship between the service quality and the customer satisfaction in the area of e-banking in Bangladesh has been carried out in [9].

The present banking situation and the scope and benefits of e-banking with compared to the existing system have been explained in [2]. Lack of secured communication is a major drawback of SMS-based mobile banking. In [10], an SMS-based banking scheme has been proposed that leverages the existing Internet based banking facility. A two-level security scheme–voice verification and digital watermarking–has been introduced.

Different e-banking services have not adequately flourished in Bangladesh. Researchers have identified some causes that are responsible for this slow deployment of e-banking. In the following we summarize the key findings identified in the existing work:

1) A reliable telecommunication infrastructure has not been built nation-wide. Specially, high-speed Internet connection is not available in every corner of the country.
2) Bangladesh need many IT professionals with experience in building and deploying e-banking services.
3) Since e-commerce and e-banking were not widely deployed before, new legislations and regulations need to be established and government policies need to be revised.
4) Due to lack of awareness, customers are not well informed about the benefits of e-banking.
5) The e-commerce and e-banking services should be free of any security breaches.

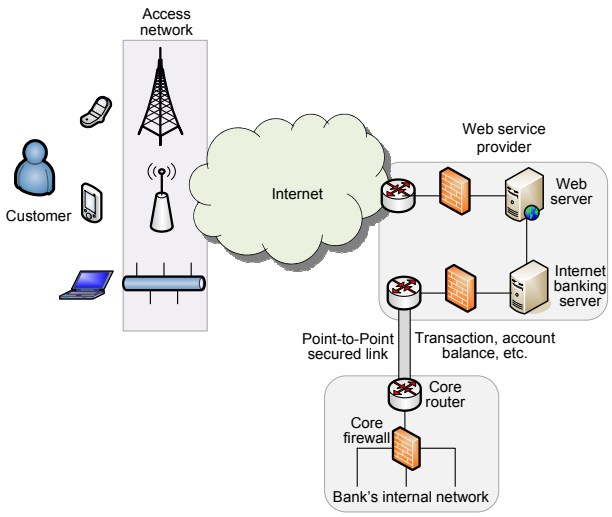


Fig. 2. The Internet banking architecture

III. Banking Network Architecture

Similar to other networked institutions, a banking network should be equipped for communications in two frontiers: communications that is needed between internal nodes and interactions with external nodes. Figure 1 shows the architecture of a typical banking network that deploys different nodes that are needed to provide e-banking services.

For the electronic transactions that are limited within the bank, the internal network is protected by a core firewall that is connected with the core router. The server zone deploys different application servers, database servers, storages, network analysis tools (e.g., SNAP). The DMZ (demilitarized zone) deploys a number of servers that are being invoked by external entities through the Internet. In general practice, sensitive information and bank's internal data are never deployed inside the DMZ.

ATM machines are established inside the branch offices and also at external premises. A branch office and the server zone is connected through a private, secured point-to-point connection, which is shown as the data link provider network between the branch router and the core router in Figure 1. The ATMs established inside a branch is connected with the server zone through this secured point-to-point connection.

The Internet banking architecture is shown in Figure 2. Internet banking customer may access the bank's Web site using any device that is able to connect to the Internet through the access network. At the Web service provider's network a Web server hosts the Web site of the bank. The customer access the Web site using one of the Web browsers. The Web service provider's network is connected with the bank's network through a secured, point-to-point connection. A third-party may provide this connection. Through this secured, private connection the Internet banking server (that resides between the Web server and the bank's internal network) transfer sensitive information that perform transaction, account balance, bill payment, etc. Hence, the customer's account related information is never being transferred through the public Internet and the security of the customer's account is being maintained.

IV. Present e-banking services in Bangladesh

E-banking might be implemented in the three main categories of e-commerce business models: Business-to-Consumer (B2C) (e.g., a customer is withdrawing cash from her bank), Business-to-Business (B2B) (e.g., funds being transferred from one bank to another) and Consumer-to-Consumer (B2C) (e.g., funds being transferred between two customers' account). In this paper, we mainly focus four types of e-banking services from the B2C business model category: ATM-based banking, tele-banking, SMS-based banking and Internet banking.

A. ATM-based banking

ATM is a virtual teller point of a bank that performs most of the tasks of a teller, including cash deposit/withdraw, balance inquiry, show statement, etc. In Bangladesh, some multinational private banks incepted the ATM booth in Dhaka in 1992. In Bangladesh, for six years from 2004 to 2010, the indicator number of ATMs (per 100,000 adults) increased from 0.14 to 1.93, which implies that the number of ATMs is growing significantly [11].

B. Tele-banking

To access tele-banking services, customers have to dial a particular telephone number provided by the bank. A customer's identity is verified by checking a PIN or security questions. The full service might be automated although an operator might be reached sometimes. A number of banking services could be realized through tele-banking including detailed account information, balance inquiry, information about products or services, ATM card activation, cheque book related services, bill payment, credit card services, etc.

C. SMS-based banking

SMS banking is a type of mobile banking, a technology-enabled service offering from banks to its customers, permitting them to operate selected banking services over their mobile phones using SMS messaging. SMS banking services are operated using both push and pull messages. Push messages are those that the bank chooses to send out to a customer's mobile phone, without the customer initiating a request for the information. Typically push messages could be either Mobile marketing messages or messages alerting an event which happens in the customer's bank account, such as a large withdrawal of funds from the ATM or a large payment using the customer's credit card, etc.

D. Internet banking

Internet banking started by launching Web sites of the banks to provide different banking related information (e.g., different types of accounts, deposit schemes, interest rate, foreign exchange rate, location of nearest branches, etc.) to the customers. Soon after its inception, the customers' accounts were integrated and thus could be operated after proper identity authentication of the customers. Internet banking if implemented in full scale would be highly appreciated by both customers and bank authorities. It would provide hassle free, 24 hours, secured e-banking services from anywhere with minimum cost. A number of banking services including account balance enquiry, fund transfer, opening or modifying term deposit account, cheque book or pay order request, exchange rate or interest rate enquiry, bill payment, etc. could be realized through Internet banking. In Bangladesh, Dutch-Bangla Bank Limited first started Internet banking in 2003.

E. Available services

At present around 50 public and private banks are operating in Bangladesh. Not all of them provide wide range of e-banking services. Table I summarizes the popular e-banking services offered by the major banks of Bangladesh. Note that the information provided in this table have been collected from various sources: yearly reports published by different banks, browsing the Web site of the banks and through personal communications with the employees of the banks. Due to limitation of space, in our list, we have included 15 banks that have large establishment. Here, A, T, S and I represent ATM, tele-banking, SMS-banking and Internet banking respectively.

TABLE I
AVAILABLE E-BANKING SERVICES OFFERED BY BANKS OF BANGLADESH

Bank	Account balance	Cash withdraw	Mini statement	Full statement	Utility payment	Fund transfer	Other services
AB Bank	[A][S][I]	[A]	[A][S]	[I]	[I]	[I]	[I]
Bank Asia	[A][S][T][I]	[A]	[S]	[I]	[I]	[I]	[I][T]
Brac Bank	[A][S][T][I]	[A]	[A][S]	[I]	[I]	[I]	[I]
Dhaka Bank	[A][S][T][I]	[A]	[A]	[I]	[I]	[I]	[I]
Dutch Bangla Bank	[A][S][T][I]	[A]	[A][S]	[I]	[I][T]	[I]	[I][S][T]
Eastern Bank	[A][S][T][I]	[A]	[A][S]	[I]	[I]	[I]	[I][T]
HSBC Bank	[A][S][T][I]	[A]	[A][S]	[I]	[I][T]	[I]	[I][S][T]
Islamic Bank	[A][I]	[A]	[A][S]	[I]	[I]	[I]	[I][S]
*Janata Bank	[A]	[A]	[X]	[X]	[X]	[X]	[X]
Mercantile Bank	[A][S][I]	[A]	[A][S]	[I]	[I]	[I]	[I][S]
Prime Bank	[A][S][I]	[A]	[A][S]	[I]	[I]	[I]	[I]
*Sonali Bank	[A]	[A]	[X]	[X]	[X]	[X]	[X]
Standard Cha-rtered Bank	[[A][S][T][I]	[A]	[A][S]	[I]	[I][T]	[I]	[I][S][T]
Trust Bank	[A][S][I]	[A]	[A][S]	[I]	[I]	[I]	[I][S]
*Uttara Bank	[A]	[A]	[X]	[X]	[X]	[X]	[X]

* Public banks operated by Bangladesh Government.

A "X" sign represents the bank has not started that service through e-banking at this moment.

Table I clearly shows that the government owned and operated public banks are far behind with compared to the private banks in implementing different e-banking services for their customers. Another important observation is Internet based banking has been implemented most among different types of e-banking services.

V. SURVEY DATA ANALYSIS

Using Google Docs [12] the present study carried out an online survey that was designed to find out the answers of the following two questions:

1) Adoption of existing e-banking services.
2) Reasons behind the present trends.

Around 75 responders participated in the survey who use Dhaka based different public and private banks. The survey was composed of two sections: the first section collects the responders' background information (e.g., gender, age, education level, etc.) and the next section accumulates the adoption of e-banking services in Bangladesh. In the following we present the survey results with our analysis.

A. Background information of the responders

Among the responders who completed the survey on-line, more than 70% were male. This is quite understandable while the number of female Internet users are much less than male users in Bangladesh. Moreover, it has been observed that young people from the age group 30-40 adopt the use of Internet more rapidly. The survey also gathered information about how frequently and for how long the responders use the Internet. Figure 3 reveals that more than 50% responders spend in the Internet more than 2 hours everyday. Only 4% responders hardly use the Internet in their daily life.

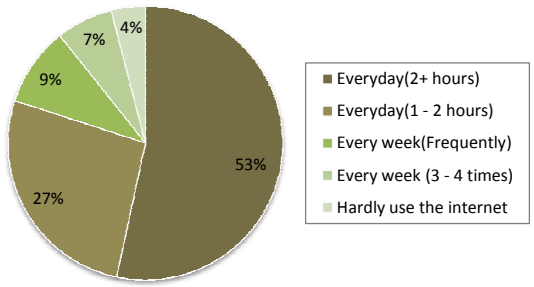


Fig. 3. Frequency of the use of the Internet by the respondents

B. Adoption of existing e-banking services

In this section of the survey, since the responders were allowed to chose more than one answers if they prefer, (in some questions) total number of responses might be higher than the number of responders (which is 70).

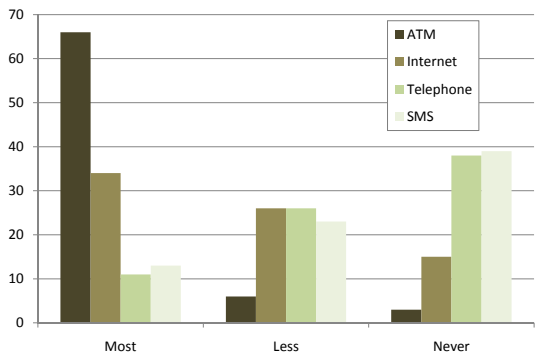


Fig. 4. Relative popularity of different types of e-banking technologies

Figure 4 summarizes the relative popularity of the four e-banking technologies that we are focusing in our study.

At present, ATM the oldest e-banking technology is the most popular among the users. After ATM Internet banking is mostly used although the difference between the usage frequency of ATM and Internet is significant. The other two technologies–tele-banking and mobile banking–could not attract many customers and are hardly used by them.

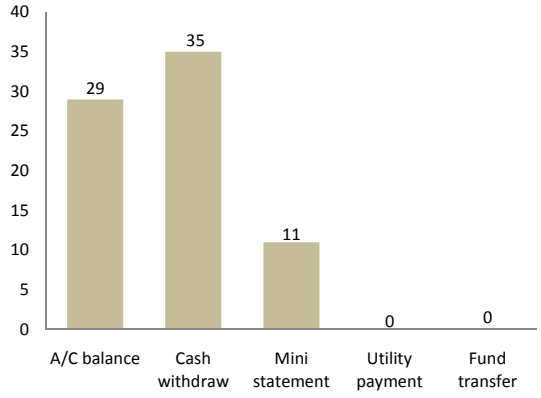


Fig. 5. Use of ATMs in different banking services

Next, our study looks at which e-banking services are mostly accessed though ATM booths. Figure 5 shows that cash withdrawal and account balance inquiry are two services that customers mostly avail though ATM booths. Although hardly used by the customers ATM could be used for utility bill payment and transferring funds. Note that these services are not provided by most of the banks through ATMs.

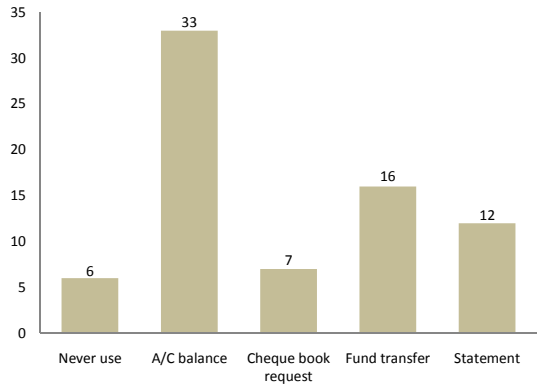


Fig. 6. Use of Internet banking in different banking services

According to Figure 6 account balance inquiry is the most popular service that customers use through Internet banking. Other than that fund transfer (within accounts of the customer herself under the same bank), view account statement and cheque book request are some useful services accessed through Internet banking.

C. Reasons behind the present trends

Information presented so far establish that at present, ATM is the only widely used e-banking technology although Internet banking is the emerging technology with great potential. Our next question to the responders was why ATM is widely used by the customers. Figure 7 shows that some of the reasons why ATM is being used widely are availability, no additional training is required and easy to use. Many ATM booths have been established all over Dhaka city and thus the banks have brought the ATM service very close to the users. People are using these machines for many years while the user interface is simple and they do not need any other training (e.g., computer or Internet knowledge) to operate these machines.

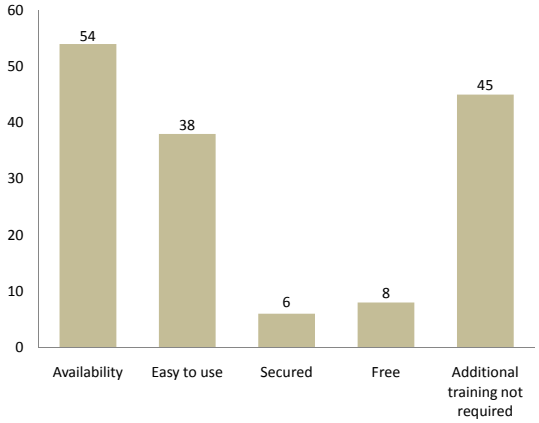


Fig. 7. Why ATM is widely used

ATM is the first e-banking technology that has been initiated in Bangladesh more than 20 years ago. On the other hand, other e-banking technologies have been available for more than 10 years. However, the other technologies could not attract the customers as the ATM did. The survey investigated the reason(s) behind this failure. Figure 8 shows the major hindrance these technologies is facing is users are not well informed.

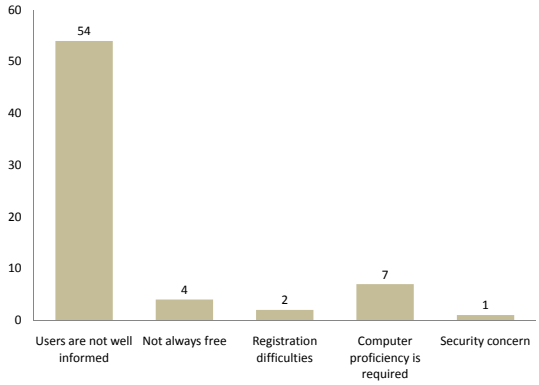


Fig. 8. Why e-banking technologies (other than ATM) are not widely used

VI. Discussion and Recommendation

Among the four types of e-banking technologies, ATM is the oldest and most mature one. SMS and tele-banking are not widely deployed. SMS has inherent security problem and its use is very limited due to unfriendly user interface. The use of tele-banking is very limited only to those automated services that do not need human assistance. Otherwise, long

waiting time in queue, poor customer satisfaction and other problems fail to make tele-banking attractive to the customers in Bangladesh. Figure 4 also establishes the fact that after ATM, Internet banking is the most widely used e-banking technology. Previously we mentioned that it is also the cheapest way of e-banking [4]. Therefore, here we discusses how the use of Internet banking might be increased.

Internet banking has great future in Bangladesh. Most of the the banks are providing at least some basic services through their Web site. The telecommunication infrastructure has significantly improved although not compared to the developed countries. Moreover, in last ten years, the number of Internet users has increased exponentially. Figure 9 shows that the number of Internet users in Bangladesh was 0.2 (per 100 people) in 2004. Only in six years, the number increased to 3.70 (per 100 people) in 2010. Figure 3 shows that 50% people spend more than two hours in the Internet. Our banks fail to bring Internet banking closer to this huge number of Internet users. Our survey also establishes the fact that e-banking or Internet banking is not widely used because the customers are not well-informed (see Figure 8).

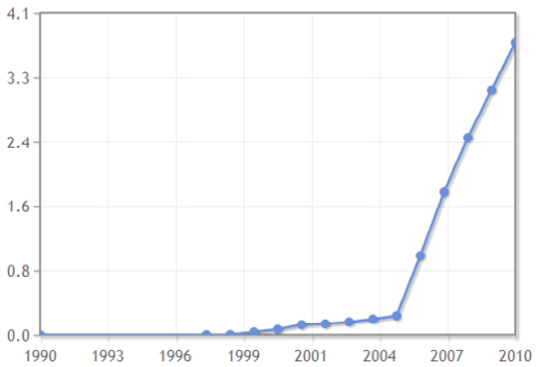


Fig. 9. Number of Internet users (per 100 people) in Bangladesh [11]

The following recommendations will make the Internet banking a cost-effective tool for the banks and the customers as well:

- Inform the customers about the conveniences of Internet banking when they open an account. Most of the customers are not aware of how easy to operate Internet banking. The customers should be demonstrated and given hands-on training.
- The Web site of the Internet banking should be easy to navigate. The log-in and other activities should be similar to other user-friendly Web sites. Moreover, considering the slow Internet speed, the Web site should be lite-weight so that it would upload quickly.
- Since customers should have trust that their accounts are safe and free from hacking, the security of the Internet banking should be given the highest attention [13].
- A significant number of users access the Internet through their cell phones and other hand-held devices, which are equipped with small screen. The banks should develop separate mobile Web sites. Moreover, mobile apps might be built to provide quick, easy, secure access to customers' bank accounts.
- The central bank (i.e., Bangladesh Bank) should update the legislations to allow fund transfer between two banks. This will increase the use of Internet banking.
- Provide more services through Internet banking including opening new accounts, maintaining investment, mortgage, buying/selling stocks in share market, transferring funds to/from foreign banks, etc. Internet banking should be a hassle-free, single entry point for all financial activities for the customers.

VII. Conclusion

The prospect of Internet banking is very good in Bangladesh while the number of Internet users are growing geometrically and most of the banks are investing in e-banking as well. This is the right time for the banks to introduce their e-banking services to their customers. The recommendations provided in the previous section could be used as guideline to develop step-by-step e-banking facilities. Although not included in this paper, mobile banking (a new trend of banking where an account is being fully operated through the customer's cell phone without any intervene with a branch) is an emerging banking technology in Bangladesh.

References

[1] A. Schaechter, "Issues in Electronic Banking: An Overview," IMF Policy Discussion Paper, No. 02/6, 2006. Available: http://www.imf.org/external/pubs/ft/pdp/2002/pdp06.pdf.

[2] M. M. Islam, "Proposed ICT infrastructure for E-banking in Bangladesh," Master of Science Thesis, Royal Institute of Technology (KTH), Stockholm, Sweden, 2005.

[3] M. Tan and T. Teo, "Factors Influencing the Adoption of Internet Banking," *Journal of the AIS*, vol. 1, no. 5, 2000.

[4] S. M. Nsouli and A. Schaechter, "Challenges of the "E-Banking Revolution"," *Finance & Development Magazine*, vol. 39, no. 3, 2002.

[5] R. R. Nitsure, "E-banking: Challenges and Opportunities," *Economic and Political Weekly*, vol. 38, 2003.

[6] Syed Shah Alam, Ali Khatibi, A. Solucis Santhapparaj and Mohammad Talha, "Development and prospects of Internet banking in Bangladesh," *Competitiveness Review: An International Business Journal incorporating Journal of Global Competitiveness*, vol. 17, pp. 56–66, 2007.

[7] Yang, J. and Ahmed, K.T., "Recent trends and developments in e-banking in an underdeveloped nation an empirical study," *Int. J. Electronic Finance*, vol. 3, pp. 115–132, 2009.

[8] A. H. M. Saidul Hasan1, Md. Azizul Baten, Anton Abdulbasah Kamil and Sanjida Parveen, "Adoption of e-banking in Bangladesh: An exploratory study," *African Journal of Business Management*, vol. 4, pp. 2718–2727, 2010.

[9] Jannatul Mawa Nupur, "E-Banking and Customers Satisfaction in Bangladesh: An Analysis," *International Review of Business Research Papers*, vol. 6, no. 4, pp. 145–156, 2010.

[10] Syed Bahauddin Alam, Hussain Mohammed Dipu Kabir, Md. Nazmus Sakib, A B M Rafi Sazzad, Celia Shahnaz and Shaikh Anowarul Fattah, "A secured electronic transaction scheme for mobile banking in bangladesh incorporating digital watermarking," in *Proc. of the International Conference on Information Theory and Information Security*, 2010, pp. 98–102.

[11] Index Mundi, "Facts about bangladesh," Available: http://www.indexmundi.com/facts/bangladesh.

[12] "Google docs," Available: http://docs.google.com.

[13] Shumaila Y. Yousafzai, John G. Pallister and Gordon R. Foxall, "A proposed model of e-trust for electronic banking," *Technovation*, vol. 23, pp. 847–860, 2003.

Experimental Framework for Mel-Scaled LP based Bangla Speech Recognition

Umme Muslima
Dept. of Electronics and Communication Engineering
University of Information Technology and Sciences
Dhaka, Bangladesh
muslima_ece@yahoo.com

M. Babul Islam
Dept. of Applied Physics and Electronic Engineering
University of Rajshahi
Rajshahi 6205, Bangladesh
babul.apee@ru.ac.bd

***Abstract*— This paper deals with the recognition process of Bangla speech. The used database consists of two sets of data – one is for training containing 3824 utterances of Bangla digit sequences of 25 male and 25 female speakers and the other one is test dataset containing 1985 utterances of 26 male and 26 female speakers. The test set is subdivided into four groups such as clean1, clean2, clean3 and clean4. Mel-LPC based front-end has been used to design the front-end, since it incorporate auditory-like frequency resolution. The Mel-LPC is a time-domain feature and computationally efficient. The Mel-LPC based cepstral coefficients are obtained directly from the input speech by using generalized autocorrelation function. In this estimation process bilinear transformation is not required, and frequency warping is obtained by using a first-order all-pass filter instead of unit delay. A detail experimental framework both for front-end and HMM based back-end have been presented in this paper. The final recognition experiments show the satisfactory performance of the developed system. The recognition accuracy are found to be 98.11%, 98.05%, 97.94%, and 97.63%, for test sets clean1, clean2, clean3 and clean4, respectively.**

***Index Terms*—Mel-LPC, HMM, Bilinear transformation, Bangla speech recognition.**

I. Introduction

Automatic recognition of spoken language is the first and probably the most important step toward natural human-machine interaction. Currently automatic speech recognition systems (ASRs) have been used widely in real world applications, for instance, in telecommunications, in question and answer (QA) systems, in car navigation systems, in robotics and so on. Research in this fascinating field in the past few decades has produced remarkable results, leading to many exciting expectations as well as new challenges in many languages, for instance, English, Spanish, Japanese, etc. [1], [2], [3], [4]. But progress of such research with Bangla language is still on primary level. Some of them could be found in [5], [6], [7]. However, Bangla is an important language with a rich heritage and is spoken by about 3.05% of the world population [8].

The speech analysis and speech representation are challenging because of the variability occurring in speech signals, the limitations of the current mathematical frameworks (e.g. Fourier transform and linear prediction) to deal with this variability. Since the human ear resolves frequencies non-linearly across the speech spectrum, designing a front-end incorporating auditory-like frequency resolution improves recognition accuracy. In nonparametric spectral analysis, Mel-frequency Cepstral Coefficient (MFCC) [9] is one of the most popular spectral features in ASR. This parameter takes account of the nonlinear frequency resolution like the human ear. But in parametric spectral analysis, the linear prediction coding (LPC) analysis [10], [11] based on an all-pole model is widely used because of its computational simplicity and efficiency. While the all-pole model enhances the formant peaks as an auditory perception, other perceptually relevant characteristics are not incorporated into the model unlike MFCC. To alleviate this inconsistency between the LPC and the auditory analysis, several auditory spectra have been simulated before the all-pole modeling [12], [13], [14], [15]. Recently, as an LP-based front-end, a simple and efficient time domain technique to estimate all-pole model is proposed in [16], which is referred to as "Mel-LPC" analysis. In this method, the all-pole model has been estimated directly from the input signal without applying bilinear transformation. Hence, the prediction coefficients can be estimated without any approximation by minimizing the prediction error power at a two-fold computational cost over the standard LPC analysis. In this paper, the Mel-Scaled LP, i.e., Mel-LPC based front-end is presented for Bangla speech recognition.

II. Mel-LPC based Analysis

In order to enhance the performance in automatic speech recognition (ASR), it is important to parameterize the perceptually relevant aspects of short-term speech spectra and their dynamics in ASR front-end. The Mel-LPC analysis is one of the most simple and efficient time domain techniques to estimate the warped all-model. It incorporates the auditory-like frequency resolution into an all-pole model by replacing the unit delay with a first order all-pass filter. We have used Mel-LPC based front-end for automatic recognition of Bangla speech.

The frequency warped signal $\tilde{x}[n]$ obtained by the bilinear transformation [17] of a finite length windowed signal $\tilde{x}[n]\,(n = 0,.....N-1)$ is defined by

$$\tilde{X}(\tilde{z}) = \sum_{n=0}^{\infty} \tilde{x}[n]\tilde{z}^{-n} = X(z) = \sum_{n=0}^{N-1} x[n]z^{-n} \quad (1)$$

where $\tilde{z}^{-1}$ is the first-order all-pass filter,

$$\tilde{z}^{-1} = \frac{z^{-1} - \alpha}{1 - \alpha . z^{-1}} \quad (2)$$

where $0 < \alpha < 1$ is treated as frequency warping factor.

The phase response of $\tilde{z}^{-1}$ is given by

$$\tilde{\lambda} = \lambda + 2 \cdot \tan^{-1}\left\{\frac{\alpha \sin \lambda}{1 - \alpha \cos \lambda}\right\} \quad (3)$$

This phase function determines a frequency mapping.

Now, the all-pole model on the warped frequency scale is defined as

$$\tilde{H}(\tilde{z}) = \frac{\tilde{\sigma}_e}{1 + \sum_{k=1}^{p} \tilde{a}_k \tilde{z}^{-k}} \quad (4)$$

where $\tilde{a}_k$ is the k-th mel-prediction coefficient and $\tilde{\sigma}_e^2$ is the residual energy [18].

On the basis of minimum prediction error energy for $\tilde{x}[n]$ over the infinite time span, $\tilde{a}_k$ and $\tilde{\sigma}_e$ are obtained by Durbin's algorithm from the autocorrelation coefficients $\tilde{r}[m]$ of $\tilde{x}[n]$ defined by

$$\tilde{r}[m] = \sum_{n=0}^{\infty} \tilde{x}[n]\tilde{x}[n-m] \quad (5)$$

which is referred to as mel-autocorrelation function.

The mel-autocorrelation coefficients can easily be calculated from the input speech signal $x[n]$ via the following two steps [16], [19]. First, the generalized autocorrelation coefficients are calculated as

$$\tilde{r}_\alpha[m] = \sum_{n=0}^{N-1} x[n]x_m[n] \quad (6)$$

where $x_m[n]$ is the output signal of an m-th order all pass filter $\tilde{z}^{-m}$ excited by $x_0[n] = x[n]$.

Due to the frequency warping, $\tilde{r}_\alpha[m]$ includes the frequency weighting $\tilde{W}(e^{j\tilde{\lambda}})$ defined by

$$\tilde{W}(\tilde{z}) = \frac{\sqrt{1-\alpha^2}}{1+\alpha\tilde{z}^{-1}} \quad (7)$$

which is derived from

$$\frac{d\lambda}{d\tilde{\lambda}} = \left|\tilde{W}(e^{j\tilde{\lambda}})\right|^2 \quad (8)$$

Thus, in the second step, the weighting is removed by inverse filtering in the autocorrelation domain using $\left\{\tilde{W}(\tilde{z})\tilde{W}(\tilde{z}^{-1})\right\}^{-1}$.

As feature parameters for recognition, the Mel-LP cepstral coefficients can be expressed as:

$$\log \tilde{H}(\tilde{z}) = \sum_{n=0}^{\infty} c_k \tilde{z}^{-n} \quad (9)$$

where $\{c_k\}$ are the mel-cepstral coefficients.

The mel-cepstral coefficients can also be calculated directly from mel-prediction coefficients $\{\tilde{a}_k\}$[20] using the following recursion:

$$c_k = -\tilde{a}_k - \frac{1}{k}\sum_{j=1}^{k-1}(k-j)\tilde{a}_k c_{k-j} \quad (10)$$

It should be noted that the number of cepstral coefficients need not be the same as the number of prediction coefficients.

III. Evaluation on Bangla Database

A. Preparation of Bangla Database

The database used in this research work is a speaker-independent Bangla digit sequences. This database consists of two sets – one is training dataset and the other one is test dataset. The number of speakers for training is 25 male and 25 female and for test datasets 26 male and 26 female and their age ranges are 19-25 years.

The database is constructed as the format of Aurora-2 [2] with some differences as shown in Table I. Table II shows the pronunciation of eleven digits in Aurora-2 and Bangla database.

TABLE I. A Comparison between Aurora-2 and Bangla Speech Database

Dataset		AURORA-2	Bangla Database
Training	No. of speakers	Male: 55 Female: 55	Male: 25 Female:25
	SNR	Clean multi-condition (clean & noisy)	clean
	Filter	G.712	--
Test	No. of speakers	Male: 52 Female: 52	Male: 26 Female: 26
	SNR	clean, 20, 15, 10, 5, 0 & -5 dB	clean
	Filter	G.712 & MIRS	--

TABLE II. Pronunciation of Bangla Digits in Bangla Database

Digit	AURORA-2	Bangla Pronunciation
1	One	/ak/
2	Two	/dui/
3	Three	/tin/
4	Four	/chaar/
5	Five	/ panch/
6	Six	/chhoy/
7	Seven	/shaat/
8	Eight	/aat/
9	Nine	/noy/
0 (Z)	Zero	/zero/
0 (O)	Oh	/shunno/

B. Experimental set up

In this research work we have designed an HMM (Hidden Markov Model) based back-end for recognition. The back-end processing refers to the reference recognizer, which is based on the HTK (Hidden Markov Model Toolkit, version 3.4) software package. The HMM was trained on clean condition. The digits are modeled as whole word HMMs with 16 states per word and a mixture of 3 Gaussians per state using simple left-to-right models. In addition, two pause models are defined. The first one called "sil" consists of 3 states with a transition structure as shown in Fig. 1. This HMM shall model the pauses before and after the utterance. The second pause model called "sp" is used to model pauses between words. It consists of a single state as shown in Fig. 2.

The front-end is designed with a 12th order prediction model of Mel-LPC analysis. The speech signal was windowed using Hamming window of length 20 ms with 10 ms frame period. As the frequency weighting function $\widetilde{W}(e^{j\widetilde{\lambda}})$ acts like a pre-emphasis, the speech signal was not pre-emphasized. The frequency warping factor was set to 0.25. Each feature vector consists of 14 cepstral coefficients and their delta coefficients including 0th terms. Thus, the feature vector size is 28. The acceleration coefficients were not used in our system.

C. Recognition performance

To measure the recognition performance of continuous speech recognition systems either the word error rate (WER) or word accuracy rate (WAR) can be used. The WER and WAR are calculated from recognition result as follows:

$$WER = \frac{S + D + I}{N} \times 100\% \quad (11)$$

$$WAR = \frac{N - D - S - I}{N} \times 100\% \quad (12)$$

where N is the total number of words. D, S and I are deletion, substitution and insertion errors, respectively.

Pre-emphasis is not required for Mel-LP based spectral analysis. Although from the theoretical point of view, pre-emphasis is not required, we conducted a recognition experiment with a pre-emphasis factor of value 0.95 and without pre-emphasis, i.e., the value of pre-emphasis factor is 0.0. In both cases, the frequency warping factor was set to 0.30. The effect of pre-emphasis on word accuracy for four test sets is listed in Table III.

The effect of warping factor on recognition accuracy is also examined and presented graphically in Fig. 3. From Fig. 3 it has been found that most of the cases maximum accuracy obtained at the value of warping factor of 0.25.

TABLE III. THE EFFECT OF PRE-EMPHASIS ON RECOGNITION ACCURACY

Condition	clean1	clean2	clean3	clean4
With pre-emphasis (0.95)	97.99	98.27	97.91	97.72
Without pre-emphasis (0.0)	98.11	98.15	98.08	97.84

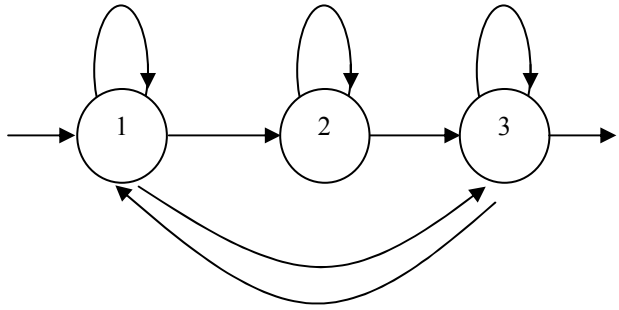


Fig. 1. Possible transition in the 3-state pause model "sil".

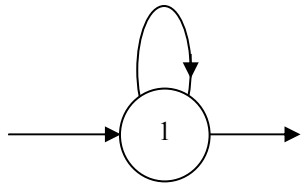


Fig. 2. Single state pause model "sp".

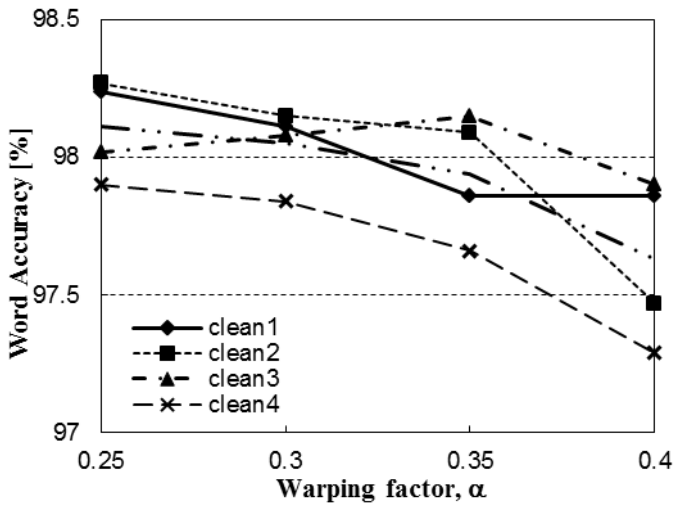


Fig. 3. Effects of warping factor on word accuracy.

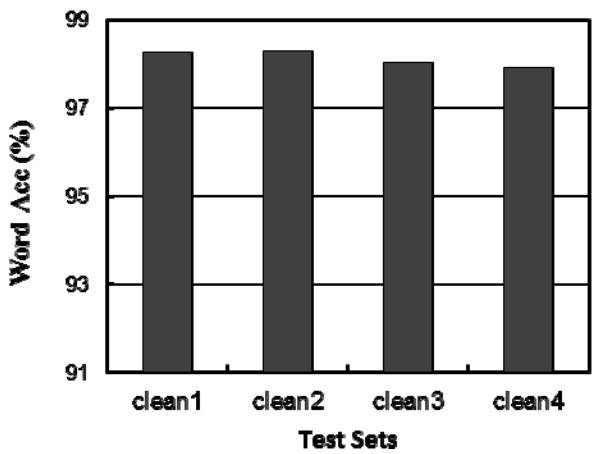


Fig. 4. The word accuracy for four test sets.

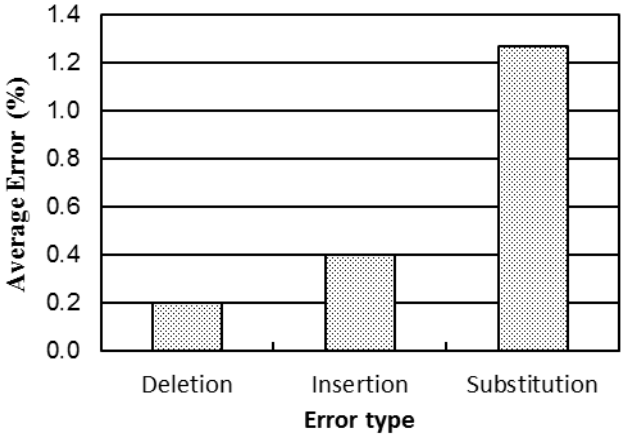


Fig. 5. Comparison among three types of errors.

The final recognition experiment was done by setting the warping factor to 0.25 and pre-emphasis factor to 0.0. The word accuracy for the four test-sets clean1, clean2, clean3 and clean4 are found to be 98.24%, 98.27%, 98.02%, 97.90% respectively. This result is presented graphically in Fig 4.

We have also examined the all three types of errors from the confusion matrix. It has been found that the deletion, insertion and subsititution errors for entire test set, i.e., clean1, clean2, clean3 and clean4 are 0.2%, 1.27%, and 0.4% on the average, respectively. A comparison of these errors has been presented in Fig. 5.

IV. Conclusion

In this paper an experimental framework for Mel-LPC based Bangla speech recognition has been proposed. The Mel-LPC spectral analysis incorporates auditory like frequency resolution into an all pole model. The remarkable point is that this analysis has been done on the linear scale without applying the bilinear transformation by using a first-order all-pass filter instead of a unit delay. In the back-end HMM based recognizer has been used.

The recognition performances for the four test sets clean1, clean2, clean3 and clean4 have been found to be 98.11%, 98.05%, 97.94%, and 97.63%, respectively. We have also examined three types of error such as deletion, insertion and substitution and average error rates have been found to be 0.2%, 0.4% and 1.27%. That is, substitution error is found to be highest on the average. This is due to the fact that there are similarities between the pronunciation of the actual and substituted word. The most confusing words are /chaar/, /panch/, /shaat/ and /aat/ . Most of the cases these words are substituted with each other.

V. References

[1] R.G. Leonard, “A database for speaker independent digit recognition”, ICASSP84, Vol.3, p.42.11, 1984.

[2] H. G. Hirsch and D. Pearce, “The AURORA experimental framework for the performance evaluation of speech recognition systems under noisy conditions”, Proc. ISCA ITRW ASR 2000: 181:188.

[3] “Spanish SDC-Aurora Database for ETSI STQ Aurora WI008 Advanced DSR Front-End Evaluation: Description and Baseline Results”, Source: Dušan Macho, UPC Version: 1.0, November 27, 2000.

[4] M. Fujimoto, K. Takeda, and S. Nakamura, “Censrec-3: An evaluation framework for Japanese speech recognition in real driving car environments”, IEIC Transaction on Information and system, vol. E89-D, no. 11: pp.2783-2793, 2006.

[5] Abul Hasanat, M. Rezaul Karim, M. Shahidur Rahman and M. Zafar Iqbal, “Recognition of Spoken letters in Bangla”, Proc. of 5th ICCIT 2002, pp. 213-216, 2002.

[6] M. Farukuzzaman Khan and R. C. Debnath, “Comparative Study of Feature Extraction Methods for Bangla Phoneme Recognition”, Proc. of 5th ICCIT 2002, pp. 27-28, 2002.

[7] G. Muhammad, Y. A. Alotaibi and M. N. Huda, “Automatic Speech Recognition for Bangia Digits”, Proc. of 12th ICCIT 2009, pp. 379-383, 2009.

[8] http://en.wikipedia.org/wiki/List_of_languages_by_number_of_native_speakers#Ethnologue_.282013.2C_17th_edition.29, viewed on 01/07/2013.

[9] S. Davis and P. Mermelstein, “Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences”. IEEE Trans. on Acoustics, Speech, and Signal Processing, ASSP-28(4): 357-366, 1980.

[10] F. Itakura and S. Saito, “Analysis synthesis telephony based upon the maximum likelihood method. Proc. of 6th International Congress on Acoustics, Tokyo: C-5-5, C17-20, 1968.

[11] B. Atal and M. Schroeder, “Predictive coding of speech signals”, Proc. of 6th International Congress on Acoustics, Tokyo: 21-28, 1968.

[12] H. Hermansky, “Perceptual linear predictive (PLP) analysis of speech”, J. Acoust. Soc. Am., 87(4): 17-29, 1987.

[13] J. Makhoul and L. Cosell, “LPCW: An LPC vocoder with linear predictive warping”, Proc. of ICASSP ’76: 446-469, 1976.

[14] S. Itahashi and S. Yokoyama, “A formant extraction method utilizing mel scale and equal loudness contour”, Speech Transmission Lab.-Quarterly Progress and Status Report (Stockholm) (4): 17-29, 1987.

[15] M. G. Rahim and B. H. Juang, “Signal bias removal by maximum likelihood estimation for robust telephone speech recognition”, IEEE Trans. on Speech and Audio Processing, 4(1): 19-30, 1996.

[16] H. Matsumoto, Y. Nakatoh and Y. Furuhata, "An efficient mel-LPC analysis method for speech recognition", Proc. of *ICSLP*98, pp. 1051-1054, 1998.

[17] A. V. Oppenheim and D. H. Johnson, “Discrete representation of signals”, IEEE Proc., 60(6): 681-691, 1972.

[18] H. W. Strube, “Linear prediction on a warped frequency scale”, J. Acoust. Soc. Am., vol. 68, no. 4, pp. 1071-1076, 1980.

[19] M. Babul Islam, K. Yamamoto and H. Matsumoto, “Mel-Wiener filter for Mel-LPC based speech recognition”, IEICE Transactions on Information and Systems, E90-D (6): 935-942, 2007.

[20] Markel, J. and A. Gray, 1976. Linear prediction of speech. Springer-Verlag.

Effect of Neural Network Based Phonetic Feature Segmentation in ASR

Mohammed Rokibul Alam Kotwal, Foyzul Hassan and Mohammad Nurul Huda
Department of Computer Science and Engineering
United International University
Dhaka, Bangladesh
rokib_kotwal@yahoo.com, foyzul.hassan@gmail.com, mnh@cse.uiu.ac.bd

***Abstract*—This paper describes a system for phone segmentation using phonetic features, where context information influences the performance of Automatic Speech Recognition (ASR). Current Hidden Markov Model (HMM) based ASR systems have solved this problem by using context-sensitive tri-phone models. However, these models need a large number of speech parameters and a large volume of speech corpus. In this paper, we propose a technique to model a dynamic process of co-articulation and embed it to ASR systems. Recurrent Neural Network (RNN) is expected to realize this dynamic process. But main problem is the slowness of RNN for training the network of large size. We introduce Distinctive Phonetic Feature (DPF) based feature extraction using a two-stage system consists of a Multi-Layer Neural Network (MLN) in the first stage and another MLN with longer context window in the second stage where the first MLN is expected to reduce the dynamics of acoustic feature pattern and the second MLN to suppress the fluctuation caused by DPF context. The experiments are carried out using Japanese triphthong and Japanese Newspaper Article Sentences (JNAS) data. The proposed DPF based feature extractor provides better segmentation performance with a reduced mixture-set of HMMs. Better context effect is achieved with less computation using MLN instead of RNN.**

Keywords*— d*istinctive phonetic feature; recurrent neural network; multi-layer neural networ;, hidden Markov model; local features

I. INTRODUCTION

Achieving Automatic Speech Recognition (ASR) at the highest possible levels of performance means making efficient use of all context information [1]. /p/ sound, for example, is different in each of "pin", "spin", and "apt". It means that specific phones that are used in any instance depend largely on context variables such as speaking rate, vocal tract length, and dialect. On a short time scale such as the average length of a phone, limitations on the rate of change of the vocal tract cause a blurring of acoustic features that is known as co-articulation. The co-articulation is especially emphasized when three vowels are connected, that is triphthong, in Japanese. For longer time scales, there are many slowly varying context variables (e.g., the degree and spectral characteristics of background noise and channel distortion) and speaker dependent characteristics (e.g., speaking rate, vocal tract length, and dialect). Hidden Markov Model (HMM) has solved these context problems by designing context-sensitive tri-phone models. But an exponential number of models are required to cover all types of context variables. Besides, acoustic models (AMs) of the HMM-based classifier include many hidden factors, such as speaker characteristics (gender, age), acoustic environments (noise), speaking style (fast, slow), etc, which affect performance in ASR.

Conventional HMM-based ASR techniques have several limitations: (a) need a large scale corpus and large numbers of mixtures to negotiate the hidden factors, (b) have degraded performance in different acoustic environments, and (c) always output an error or rejection when a new vocabulary or so-called out-of-vocabulary (OOV) word is included in an utterance. Therefore, a new approach of ASR system with high recognition performance and low computational cost is needed for practical applications. In a previous work, Distinctive Phonetic Feature (DPF) based ASR system was introduced, where Multi-Layer Neural Network (MLN) was used to extract DPFs [2]. The DPF-based system maps an input acoustic feature onto DPF space, which is invariant to acoustic environments, widens the margin of acoustic likelihood, which results in less number of mixtures in HMM and comparable recognition accuracy. However, single MLN suffers from its inability to model dynamic information.

In this paper, from the segmentation performance and computational points of view, we investigate and evaluate four types of DPF-based feature extraction methods together with the conventional method of MFCC. Four types of feature extractors are (i) MLN [2], (ii) RNN (Recurrent Neural Network), (iii) RNN-MLN (iv) two-stage MLN with smaller context window [3] and (v) the proposed, two-stage MLN with longer context window. RNN can represent dynamics in a sequence of acoustic features, that is, uncertain evidence is accumulated over many time frames in order to build up an accurate representation of the long term context variables. But the main problem of RNN is the slowness of network of large size during training. Without proper training for large network better context effect is impossible [1]. It is expected that the proposed approach generates generates more precise phonetic segment as well as higher recognition accuracy with less computation by using a MLN in the first stage for the reduction of pattern dynamics and another MLN in the second stage for the reduction of fluctuation in DPF context, and consequently

gives an accurate phonetic typewriter functionality that will assist the next generation ASR systems in solving the OOV-word problem. The experiments are carried out using Japanese triphthong data and Japanese Newspaper Article Sentences (JNAS).

The paper is organized as follows: Section II discusses about DPF-based phonetic segmentation. Section III explicates the system configuration of the existing methods. Section 4 illustrates the system configuration of RNN-MLN approach. The proposed method is explained in Section 5. Experimental database and setup are provided in Section 6, while experimental results are analyzed in Section 7. Finally, Section 8 draws some conclusion with future some remarks.

II. DPF-BASED PHONETIC SEGMENTATION

A phone can be easily identified by using its unique Distinctive Phonetic Feature (DPF) set [4], [5]. Because the traditional DPF set (high, low, anterior, back, coronal, plosive, continuant, fricative, nasal, voiced, semi-vowel) is not designed for ASR system, the feature vector space composed of the traditional DPF is not necessarily suitable for classifying speech signal [2]. A novel DPF set with 15 elements is introduced by modifying traditional set [6] [7] [8] [9]. These DPF values are mora, <high, low, nil>, <anterior, back, nil>, coronal, plosive, affricate, continuant, voiced, unvoiced, nasal and semi-vowel. Table I shows the distinguishable features of vowels. Here, present and absent elements of the DPF are indicated by "+" and "-" signs, respectively. Phonetic segmentation of utterance /iai/ for "high" and "back" DPFs are shown in Figure 1. Solid and dotted lines indicate actual and desired segmentation, respectively, while horizontal and vertical axes present respective time and DPF values.

TABLE I. DPFS FOR VOWELS

	High	Low	Nil	Back	Nil
a	-	+	-	+	-
i	+	-	-	-	+
u	+	-	-	+	-
e	-	-	+	-	+
o	-	-	+	+	-

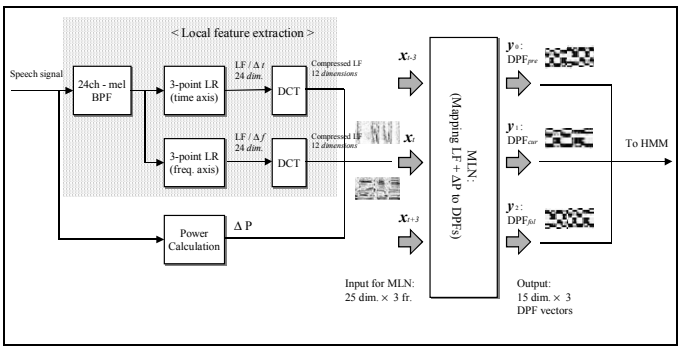


Fig. 1. System for MLN-based DPF extractor.

Conventional approach of ASR systems uses Mel Frequency Cepstral Coefficient (MFCC) as feature vector to be fed into a classifier. Parameters (mean μ and covariance Σ of hidden Markov model of each phone) are estimated from MFCC training data. Minimum Mahalanobis distance is calculated for each frame of test utterance and hence, a minimum distance phone is obtained for correct phone segmentation.

Local feature (LF) as acoustic vector was introduced in [10], and it was shown that LF was superior to MFCC as the input to MLNs [2]. T. Fukuda, et al. implemented DPF based phonetic segmentation for LF input acoustic vector using MLN [2].

A. DPF-based phonetic segmentation using MLN

Figure 2 shows a process of DPF-based phonetic segmentation using MLN. At the acoustic feature extraction stage, input speech is converted into two LFs (LF/Δ_t, LF/Δ_f), which represent a variance of spectrum along the time and frequency axis by using three-point linear regression (LR) calculation. LFs are input to an MLN with three layers, including 2 hidden layers, after combining preceding t_{i-3} and succeeding t_{i+3} frames with the current frame t_i. Each frame of LF consists of 25 values (12Δt+12Δf+ΔP). The MLN has 45 output units (15x3) corresponding to a set of triphthongs (three consecutive vowels), or context- dependent DPF vector, that consists of three DPF vectors (a preceding context DPF, a current DPF, and a following context DPF) with 15 dimensions each. The two hidden layers consist of 150 and 90 units from the input layer. Because these types of neural networks (NN) have no feedback connection, it is difficult to embed contextual information into them. The MLN is trained by using a standard back-propagation algorithm.

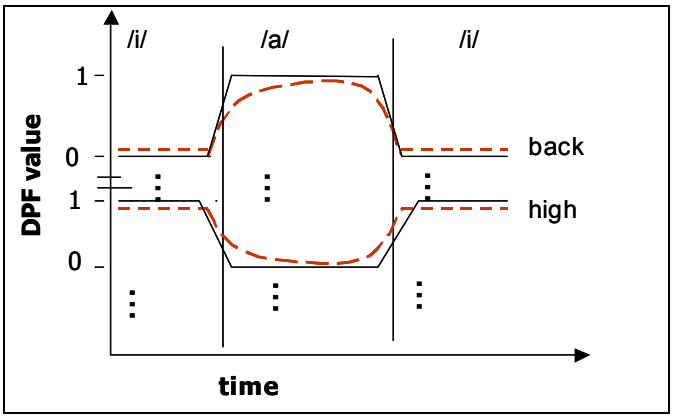


Fig. 2. DPF-Based Phonetic Segmentation.

B. DPF-based segmentation using RNN

RNN for its capability of handling a longer context window is the basics of this network. The aim of the network is to perform the mapping from a sequence of frames of parameterized speech to a sequence of phone labels associated with those frames as well as capturing contextual effects that is now a challenging task. A single frame of parameterized speech is replaced by an acoustic vector containing several adjacent frames along with the original central frame to increase the dimensionality of the acoustic vector. Figure 3

shows a block diagram of a DPF-Based phonetic segmentation method using an RNN. Higher dimensional input acoustic vector is formed by taking preceding (t_{i-3}) and succeeding (t_{i+3}) frames together with the current frame (t_i). Each input frame is formed by 25 LF values same as DPF-based phonetic segmentation using MLN. The RNN generates 45 DPF values of which 15 are for the preceding frame, 15 for the current frame and the rest for the succeeding frame. An MLN is used to integrate these 45 DPF values to obtain 15 DPF output for the current frame.

A fully recurrent neural network (FRNN) of two layers is used for this approach. Hidden layer consists of 180 hidden units. Each time total input vector is formed by taking output layer (OL) feedback values and hidden layer (HL) feedback values together with external input (25x3) LF values of that time. Feedback values of hidden and output layer at time t_0 are assumed to be 0.1. Epoch wise back-propagation through time algorithm is used for training. MLN for this approach has 3 layers (2 hidden layers, output layer). Hidden layer one and two are of 90 and 30 hidden units, respectively. Standard back-propagation is used as learning algorithm.

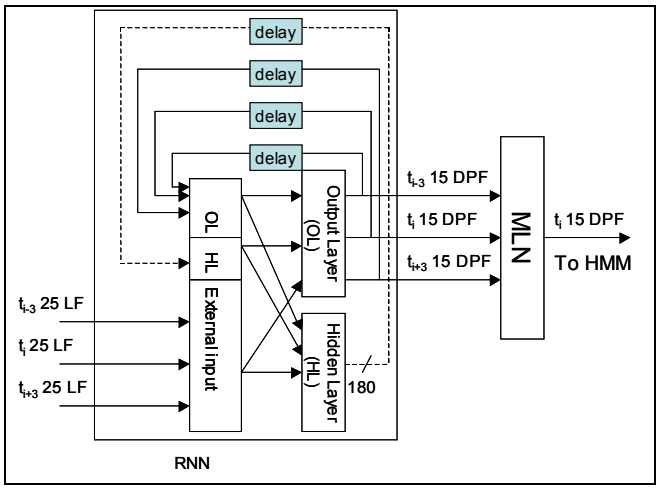


Fig. 3. Diagram of RNN-based DPF extractor.

C. DPF based segmentation method with two-stage MLN[3]

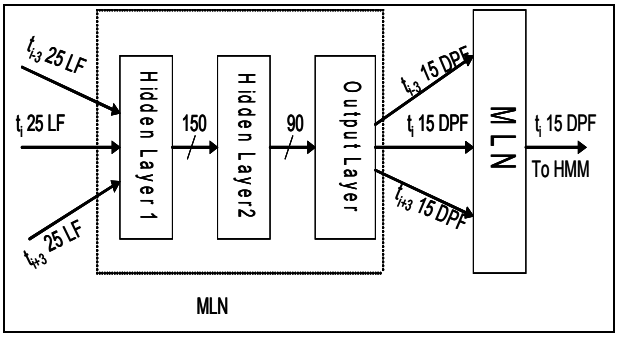


Fig. 4. System for two-stage MLN DPF extractor [3].

Figure 4 depicts the system configuration of DPF-based phonetic segmentation using two-stage MLN. Single MLN can not handle context effect rigorously. Again, DPF-based segmentation using RNN can not provide good segment and contextual information without proper training of large and complex network. Highest level performance of ASR faces challenge to capture the context information. The proposed system replaces the first stage RNN of DPF-based system with RNN described in Section 4, by a MLN to reduce the time complexity to a certain limit. Again better context effect is maintained by forming higher dimensionality of input acoustic vector of preceding (t_{i-3}) and succeeding (t_{i+3}) frames together with the current frame (t_i). Each input frame is formed by 25 LF values same as DPF-based phonetic segmentation using MLN.

The first stage MLN, named $MLN_{LF\text{-}DPF}$, generates 45 DPF values of which 15 are for the preceding frame, 15 for the current frame and the rest for the succeeding frame from LF. The second stage MLN, named MLN_{Dyn}, is used to correlate these 45 DPF values to obtain 15 DPF output for the current frame, and thereby, realizes dynamic property of utterance one degree higher than that using only one stage MLN. The MLN_{Dyn} also reduces the DPF fluctuation caused at starting of an utterance and the effect of previous phone on the current phone.

The $MLN_{LF\text{-}DPF}$ uses three layers (2 hidden layers, output layer). First hidden and second hidden layer from the input side and output layer are of 150, 90, and 45, respectively. Each time total input vector is formed by taking the external input (25x3) LF values of preceding, current and succeeding time. Standard back-propagation algorithm is used for training the network of reasonable size. The MLN_{Dyn} has 3 layers (2 hidden layers, output layer). Hidden layer one and two are of 90 and 30 hidden units, respectively. Standard back-propagation algorithm is also used for training MLN_{Dyn}.

III. THE PROPOSED METHOD

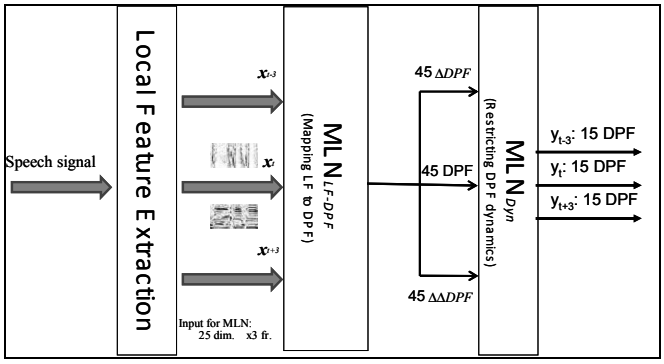


Fig. 5. The proposed method.

In this method proposed by us, two MLNs instead of a single MLN are used to construct the DPF extractor; Figure 5 shows the corresponding system diagram. The first MLN, MLNLF-DPF, maps acoustic features, or LFs, onto discrete DPF features and the second MLN, MLNDyn, reduces misclassification at phoneme boundaries by constraining the DPF context. Here, MLNLF-DPF has the same architecture as that described in subsection II. ©, and it is trained using the same learning algorithm. The 45-dimensional context-dependent DPF vector provided by the MLNLF-DPF, and its corresponding ΔDPF and ΔΔDPF vectors calculated by three-

point LR are appended into the subsequent MLNDyn with four layers including two hidden layers of 300 and 100 units, respectively. The MLNDyn is trained using the standard back-propagation algorithm and outputs a 45-dimensional DPF vector (15 for preceding context, 15 for current context, 15 for succeeding context) in which context effects for the current "t"-th frame are reduced. Since this method inserted ΔDPF and ΔΔDPF into the second stage, this method controls dynamics and reduces context effects, and hence, provides better segmentation (less deviation from ideal boundary) and shows better DPF patterns in comparison with the DPF extractor implemented by a single MLN.

IV. EXPERIMENTS

Experiments are performed by using conventional MFCC method, DPF-based method with LF+MLN, DPF-based method with RNN, DPF-based method with LF+RNN+MLN, and the proposed DPF-based method with LF+MLN+MLN. Frame correctness is calculated by measuring the distance between teacher and the frame in question. Recognition rate (%) is calculated using HMM filler (monophone) models. Figure 6 illustrates the monophone HMM filler model.

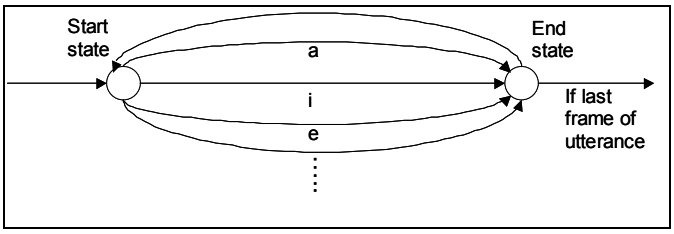


Fig. 6. Monophone HMM filler model.

A. Speech Database

A subset of "ASJ (Acoustic Society of Japan) continuous speech Database", consisting of 1316 sentences uttered by male speakers is used as training. 60 tripthongs (vowel only, such as /aei/, /uoi/, etc.) are used as test data. Sampling rate is 16KHz.

B. Experimental Setup

The Frame length and frame rate are set to be 25 ms and 10 ms, respectively. MFCC consists a vector of 38 dimensions (12 MFCC, 12Δ, 12 ΔΔ, ΔP and ΔΔP, where P is log energy of raw signal). Input features for HMMs are given below.

i) MFCC <Conventional>

ii) LF+MLN <Fukuda '04>

iii) LF+RNN

iv) LF+RNN+MLN

v) LF+MLN+MLN <Proposed>

In LF+MLN+MLN method [3], two varied dimensions are used for HMM filler model. In one variation, 15 DPF vectors are directly fed into the classifier, while in another, delta and delta-delta coefficients are appended to 15 DPFs (so, a total of 45 dimensions) before being fed into the classifier. Log normal distribution [11] [12] and Gaussian distribution are used in the first and the second case, respectively. The HMM-based classifier in this experiment adopts a standard monophone-based HMM with 5-state 3-loop left-to-right models.

Frame correct rate is calculated by the following equation.

$$\text{frame correct rate} = (1 - \frac{No.of\ erroneus\ frames}{Total\ no.of\ frames}) \times 100\%$$

V. EXPERIMENTAL RESULTS AND DISCUSSION

Table II shows frame correct rate for the conventional method MFCC, the existing method LF+MLN, LF+RNN, LF+RNN+MLN, and the proposed method LF+MLN+MLN. MFCC shows frame correct rate 54.72%, which is the worst over all the methods. The proposed method, LF+MLN+MLN shows the highest number of frame correct rate (88.16%). The existing method, LF+MLN generates frame correct rate (74.51%), which is 13.65% lower than the proposed method (88.16%). LF+RNN provides frame correct rate 70.62%, which is less than the corresponding rate of existing LF+MLN method. Since large and complex network in LF+RNN is not possible because of slowness, hence poor performance of segmentation is shown. On the other hand, if we connect extra MLN with LF+RNN, it provides better segmentation accuracy (76.01%) than the conventional method. Above discussion indicates that the proposed method can perform more precise phone segmentation than the other investigated methods.

TABLE II. FRAME CORRECT RATE (%) WITH DIFFERENT METHODS (NOT USING FILLER MODELS)

Investigated methods	Frame correct rate(%)
MFCC(38)	54.72
LF+MLN(45)	70.17
LF+MLN(15)	74.51
LF+RNN(15)	70.62
LF+RNN+MLN(15)	76.01
LF+MLN+MLN(15)	88.16

TABLE III. RECOGNITION RATE (%) WITH DIFFERENT METHODS USING HMM FILLER MODELS. THE RESULTS ARE SHOWN FOR MIXTURES 1, 2, 4, AND 8

Investigated Methods	Recognition Rate(%)			
	Mix.1	Mix.2	Mix.4	Mix.8
MFCC(12mfcc+12Δt +12ΔΔt+ ΔP+ ΔΔP)	69.57	76.81	71.01	76.81
LF+MLN(45 dpf)	68.12	72.46	75.36	79.71
LF+MLN(15 dpf)	72.46	72.46	75.36	81.16
LF+RNN(15dpf+15 Δ+15 ΔΔ)	69.57	73.91	72.46	78.26
LF+RNN+MLN(15dpf+15 Δ+15 ΔΔ)	72.46	76.81	78.26	78.26
LF+MLN+MLN(15dpf+15 Δ+15 ΔΔ)	75.36	78.26	81.16	78.26
LF+MLN+MLN(15dpf)	75.36	76.81	75.36	81.16

Table III shows the evaluated result by HMM monophone filler models. The numbers of Gaussian mixtures in HMM are used as 1, 2, 4 and 8. From the table, we can observe that MFCC again has the lowest number of recognition rate (69.57%) with mixture 1. Our proposed method provides the highest number of recognition rates (75.36%) with the same mixture.

The proposed LF+MLN+MLN achieved a significantly high performance with mixture 2 (78.26%). But LF+RNN+MLN shows 76.81% recognition rate at that mixture. No other methods in the experiment achieve comparable performance with this low number of mixture.

The highest recognition accuracy (81.16%) using Gaussian distribution at mixture 4 is obtained by using the proposed LF+MLN+MLN method with delta and delta-delta parameters. This result indicates that the proposed method provides higher number of recognition accuracy with less computation (less number of HMM mixtures).

The lower computation demand is fulfilled by the proposed method providing better recognition rate at lower number of mixtures. Increment of mixture requires k-mean clustering with large computational requirements to estimate parameters (μ^m, Σ^m) where m indicates the size of mixture.

Figure 7 shows the phonetic segmentation of utterance [iai] for the proposed LF+MLN+MLN method, LF+MLN, and LF+RNN+MLN where considering DPF value is “back”. Among all the line graphs, segmentation by LF+MLN and LF+RNN+MLN show worse performance than the proposed LF+MLN+MLN method.

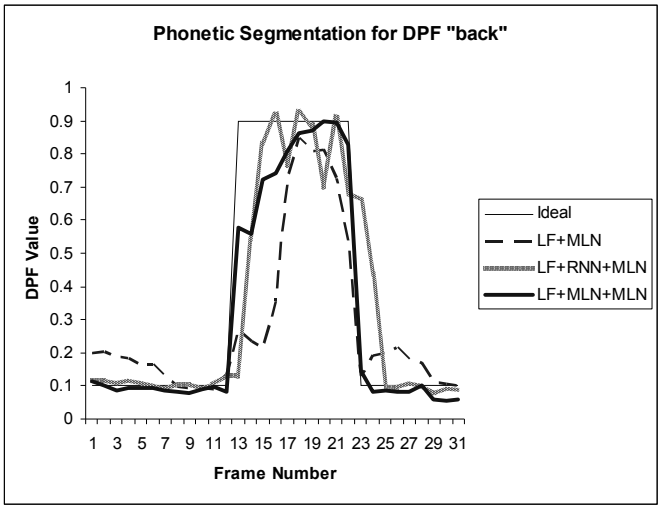


Fig. 7. Investigated Segmentation for /iai/.

LF+MLN shifts its phone boundary inward from its ideal position. A spike of DPF fluctuation is found at starting point in the existing method. These both problems of segmentation are reduced by introducing another MLN with LF+MLN in the proposed method. On the other hand, the LF+RNN+MLN method displace its phone boundary outward. The proposed method shows no displacement of boundary for this utterance.

Experimental data shows that the most of the errors occur at beginning frames in our approach. On the contrary, LF+MLN shows the most of the errors in the transition point. LF+RNN+MLN can solve better context effect than LF+MLN but it expands phone boundary up to several frames. As a result, outward displacement of phone boundary is observed in LF+RNN+MLN. The existing LF+MLN method cannot realize the short duration transitional effect. Therefore, contextual effect cannot properly solved by this network. The proposed method shows more precise segmentation with less computation. It can determine transition point because of context frame for each time of the network training. It claims that LF+MLN+MLN can solve better context effect than LF+MLN and can determine transition point with less computation than LF+RNN+MLN.

For the systems including our proposed one, a subset of the Acoustic Society of Japan (ASJ) [13] Continuous Speech Database comprising 4503 sentences uttered by 30 different male speakers (16 kHz, 16 bit) is used as training data. This test data set comprises 2379 JNAS [14] sentences uttered by 16 different male speakers (16 kHz, 16 bit).

Segmentation for a clean /jiNkoese/ utterance are shown in Figures 8, 9, 10, and 11 for a balanced-DPF set using a single MLN, MLN+MLN, RNN, and RNN+MLN, respectively. In the figures, "Solid thin line" and "Solid bold line" represent "ideal segmentation" and "output segmentation", respectively; "nasal", "nil(high/low)", and "high" of phoneme /N/, and "unvoiced", "coronal", and "anterior" of phoneme /s/ are denoted by circled 1, 2, 3, 4, 5, and 6, respectively. After observing circled 1, 2, 3, 4, 5, and 6 marked places of all the figures, we can say that the MLN+MLN exhibits more precise segmentation (less deviation from ideal boundary), reduces some fluctuations and provides more smoothed DPF curves, and hence, it misclassifies fewer phonemes.

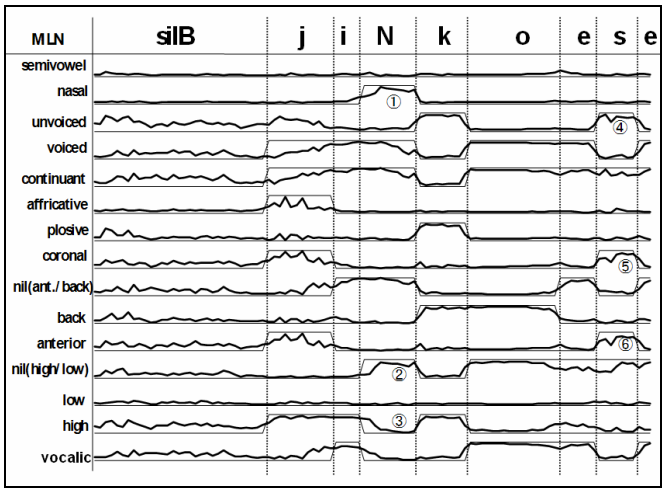


Fig. 8. Segmentation for utterance, /jiNkoese/ using MLN.

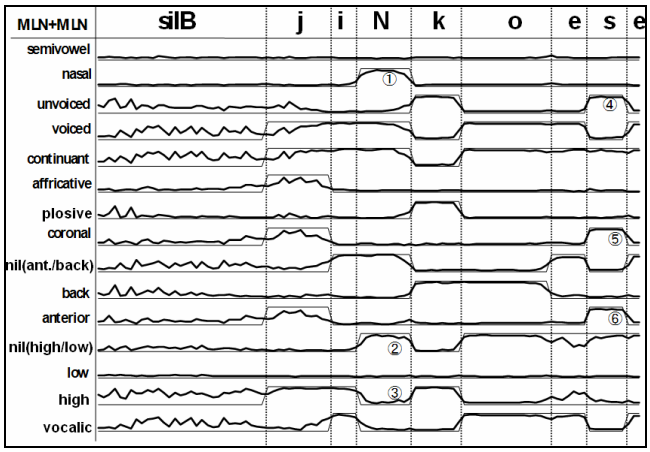


Fig. 9. Segmentation for utterance, /jiNkoese/ using MLN+MLN.

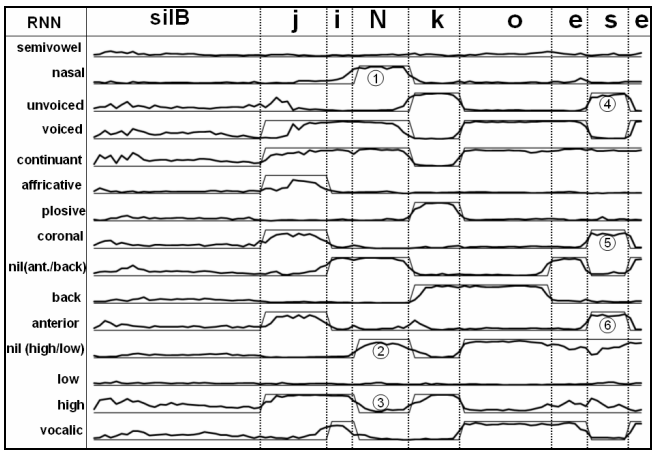


Fig. 10. Segmentation for utterance, /jiNkoese/ using RNN.

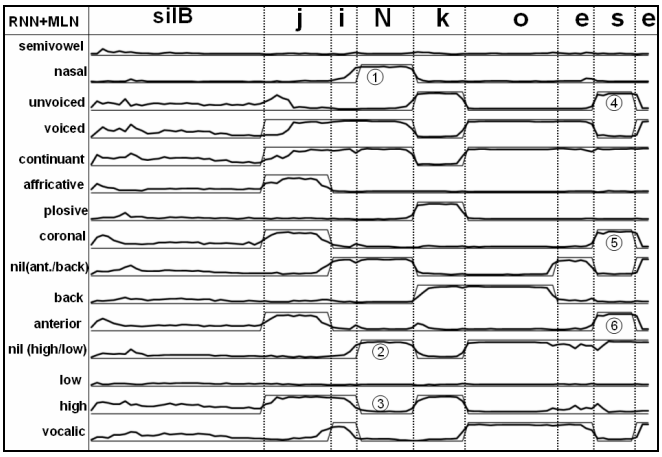


Fig. 11. Segmentation for utterance, /jiNkoese/ using HNN or RNN+MLN.

Figures 12 and 13 show the effect of extra MLN over MLN and RNN, respectively by depicting phoneme-wise DCR. In the Fig. 12, the MLN+MLN provides higher DCRs for all the phonemes except /o/ and /silB/. Because the second MLN, MLNDyn, solves coarticulation problems more widely by taking dynamic feature parameters, ΔDPF and ΔΔDPF, as input, the MLN+MLN exhibits better segmentation as well as higher DCR. Again, in the Figure 13, the RNN+MLN exhibits more accurate segmentation over RNN because extra MLN reduces context effects, and hence, gives higher DCRs for all phonemes except /my/, /dy/, /gy/, /hy/, /ry/, and /py/.

The effect of extra MLN over MLN and RNN by considering correct rate (CR) for Japanese-DPF set is shown in Figures 14 and 15, respectively. From Figure 14, the MLN+MLN shows higher correct rates for all DPFs except "vocalic", "high", and "voiced", while higher CRs for all DPFs are observed by the RNN+MLN in Figure 15.

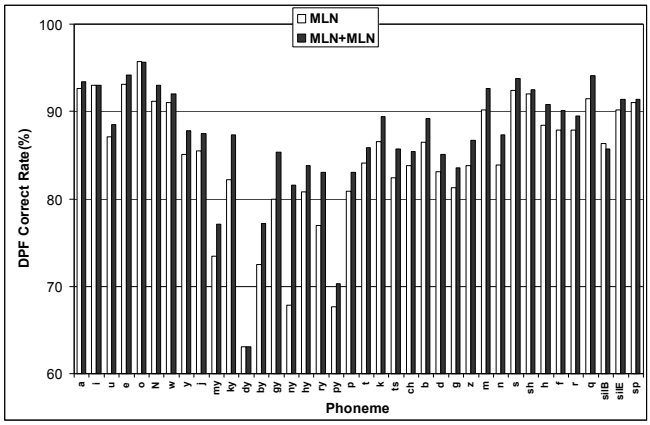


Fig. 12. DPF correct rate for each phoneme using MLN and MLN+MLN.

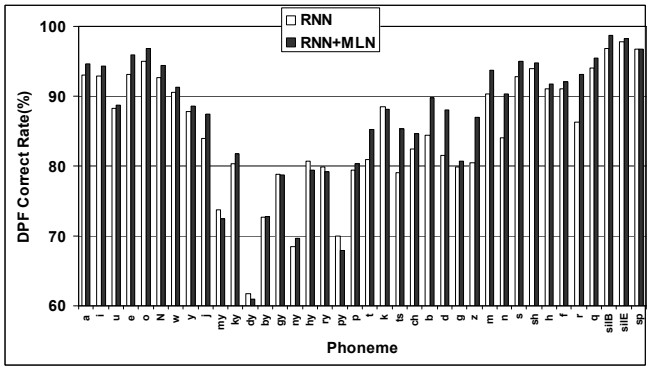


Fig. 13. DPF correct rate for each phoneme using RNN, and HNN or RNN+MLN.

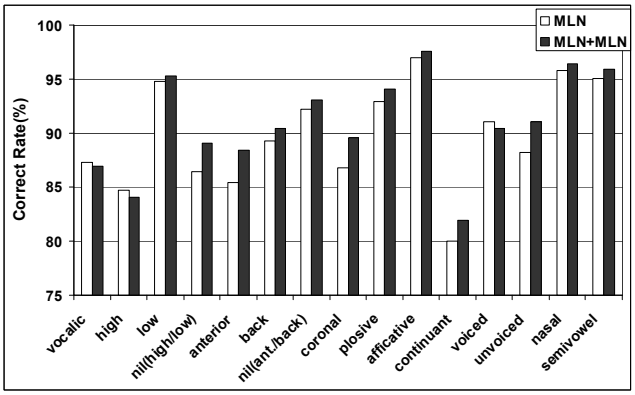


Fig. 14. Correct rates for Japanese balanced-DPF set using MLN and MLN+MLN

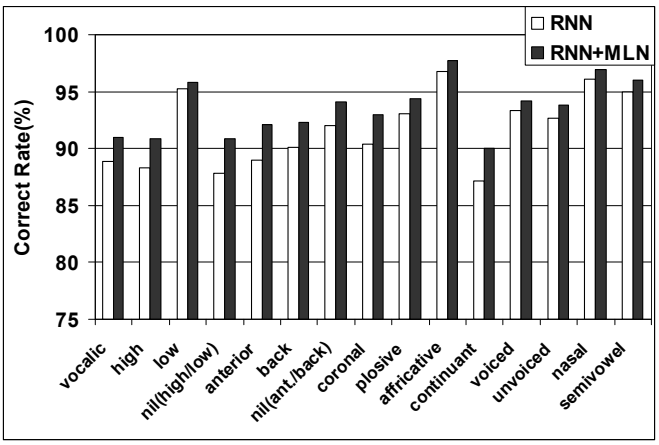


Fig. 15. Correct rates for Japanese balanced-DPF set using RNN, and HNN or RNN+MLN.

The overall DCRs for the MLN, RNN, MLN+MLN, and RNN+MLN are shown in Table IV; the MLN+MLN exhibits 1.2% improvement of DCR over the DPF extractor implemented by a single MLN, while an improvement of 1.8% is shown by the RNN+MLN over a single RNN.

TABLE IV. OVERALL DPF CORRECT RATES

DPF Extraction Method	DPF Correct Rate (%)
MLN	89.80
RNN	91.70
MLN+MLN	91.00
RNN+MLN	93.50

VI. CONCLUSION

This paper has presented a relatively simple phone segmentation system with a reduced mixture-set of HMMs by incorporating the powerful mechanism acoustic context. DPF-based method using two-stage MLN with smaller context [4] achieved accurate segmentation on triphthong task. This study concludes the following

i) The DPF-based method with smaller context [4] has also provided high recognition accuracy with less computation on triphthong data.

ii) The addition of one extra MLN over RNN or MLN provides more precise segmentation than a single MLN or a single RNN.

iii) The proposed two stage MLN with longer context shows higher DPF correct rate, phoneme correct rate and more precise segmentation over the other methods investigated.

The author would like to do further investigation by incorporating language model with the proposed method.

REFERENCES

[1] T. Robinson, "An application of Recurrent Nets to Phone Probability Estimation," IEEE Trans. Neural Networks, Volume 5, Number 3, 1994.

[2] T. Fukuda, et al, "Orthogonalized Distinctive Phonetic Feature Extraction for Noise-robust Automatic Speech Recognition," IEICE Trans. Info. & Sys., vol.E87-D, no.5, pp.1110-1118, 2004.

[3] S. King and P. Taylor, "Detection of Phonological Features in Continuous Speech using Neural Networks," Computer Speech and Language 14(4), pp. 333-345, 2000.

[4] E. Eide, "Distinctive Features for Use in an Automatic Speech Recognition System," Proc. Eurospeech 2001, vol.III, pp.1613-1616, 2001.

[5] T. Fukuda and T. Nitta, "A Study on Japanese Distinctive Phonetic Feature Set for Robust Speech Recognition," The 2003 Autumn Meeting of The Acoustical Society of Japan, Vol.I, 1-6-5, pp.9-10, September 2003, *in Japanese*.

[6] T. Fukuda and T. Nitta, "Noise-robust Automatic Speech Recognition Using Orthogonalized Distinctive Phonetic Feature Vectors," IPSJ SIG Technical Reports, 2003-SLP-47 (15), pp.77-82, July 2003, *in Japanese*.

[7] T. Fukuda, W. Yamamoto and T. Nitta, "A Study on Robust Speech Recognition by Using Distinctive Phonetic Feature Vectors," IEICE Technical Report, SP2002-121, pp.1-6, December 2002, *in Japanese*.

[8] T. Fukuda, W. Yamamoto and T. Nitta, "A Study on an Automatic Speech Recognition by Using Distinctive Feature Vectors," The 2002 Autumn Meeting of The Acoustical Society of Japan, Vol.I, 1-9-1, pp.1-2, September 2002, *in Japanese*.

[9] Nitta, et al, "Representing local features for speech recognition in cepstrum domain," Proc. Spring Meeting of ASJ, pp.131-132, 2001.

[10] T. Fukuda and T. Nitta, "Automatic Speech Recognition by Using Distinctive Phonetic Features Approximated with Logarithmic Normal Distribution," IEICE Technical Report, SP2003-23, pp.19-24, May 2003, *in Japanese.*

[11] T. Fukuda, W. Yamamoto and T. Nitta, "A Study on Robust Speech Recognition by Using Logarithmic Normal Distribution of Distinctive Phonetic Feature Vectors," The 2003 Spring Meeting of The Acoustical Society of Japan, Vol.I, 3-Q-2, pp.155-156, March 2003, in Japanese.

[12] ASJ: The Acoustic Society of Japan. http://www.asj.gr.jp/eng/index.html

[13] JNAS: Japanese Newspaper Article Sentences. http://www.milab.is.tsukuba.ac.jp/jnas/instruct.html.

A Framework for Bangla Text to Speech Synthesis

K. M. Azharul Hasan, Muhammad Hozaifa
Computer Science and Engineering Department
Khulna University of Engineering & Technology
Khulna 9203, Bangladesh.
azhasan@gmail.com, hozaifa.moaj@yahoo.com

Sanjoy Dutta, Rafsan Zani Rabbi
Computer Science and Engineering Department
Khulna University of Engineering & Technology
Khulna 9203, Bangladesh.
dsanjoy58@live.com, rzrt.edu@gmail.com

***Abstract*— we describe a basic framework and methodology to convert Bangla Text to Speech. Articulated words are automatically produced from Bangla input text by the methodology from the basic pronunciation of the Bangla words. The single tone syllables are considered as the fundamental units for analysis. The methodology selects phonetic units from uttered vocabulary and then combined the appropriate diphones to get the final output. The uttered syllables are selected on the basis of synthesized input text. The input text is analyzed according to the very basic grammar rules of Bangla pronunciation. Using the methodology, we developed a prototype system and the output generated from the system has been analyzed with the natural human voice**

***Index Terms*— Bangla Text to Speech, Speech synthesis, Text Normalization, diphone preparation Syllable Parser.**

I. Introduction

Text-to-speech (TTS) is a software or hardware system for speech synthesis application that is used to create a spoken sound version of the text from a computer document. The purpose of TTS is to enable the reading of computer document or information for the visually challenged person, or may be used to augment the reading of a text message[1]. Text-to-speech synthesis is a complex combination of language processing and signal processing in computer science[2]. Speech synthesis is the artificial production of human speech. Text to speech system consists of two main component, text analysis and speech generation [3]. In the text analysis part, text are processed and parsed on the basis of the pronunciation rules of the language and appropriate sequences of phonemic units are produced. Speech generation component uses these phonemic units to select the respective syllable and generate speech. Accuracy of the speech depends on the text analysis component[3][4]. Bangla is one of the important Indo-Iranian languages which is the sixth-most popular in the world and spoken by a population that now exceeds 250 million. It is the primary language in Bangladesh and second language in India[5]. Text to speech synthesis has a long history and lots of research has been done on different languages such as English[1][6], French [7][8] Chinese [9] etc. But in contrary, text to speech analyses for Bangla language still an unsolved research problem due to lake of resources such as speech corpus, labeled data etc [10]. Text to speech synthesis is often done by concatenation of natural speech. The natural speech are concatenated together to form a word or sentence. Concatenated speech synthesis has become very popular in recent years due to its improved sensitivity of the output[11].

In this paper, we propose a formal framework for text to speech synthesis for Bangla. The framework is based on the basic pronunciation techniques of Bangla grammar. We used different parts of each letter as the basic unit of speech and then concatenated the units to get the word or sentence of the TTS. We also synthesize the Bangla words for pronunciation based on Bangla grammar. The proposed system will be highly useful and necessary for an illiterate or vision impaired Bangla speaking people to hear and understand the content of a document where they face many problems in their daily life due to the differences in their script system.

II. Related Works

There are some works for Bangla TTS namely [14]-[18] that handles Bangla text to speech for different perspective. Speech is synthesized by considering the sequential flow of diphthongs in [14]. This is based on the idea of Bangla consonant phoneme that is overlapped by the modifier phonemes and produces a single diphthong. [15] presents the design and development of unrestricted TTS system for Bangla language. Unrestricted TTS system can synthesize good quality of speech in different domains. A methodology for text normalization and diphone preparation for Bangla TTS is presented in [16] on the basis of Concatenation based system. [17] describes an implementation of a Bengali speech synthesizer for mobile devices. Epoch Synchronous Non Overlap Add (ESNOLA) is used for speech generation. The scheme is based on concatenated speech synthesis technique and part names are used as the smallest signal units for concatenations. A Bangla TTS system was developed using open source Festival in [18]. This system is developed using diphone concatenation approach in its waveform generation phase and shows some practical applications of it. All the works [15]-[18] deals with different forms or parts of Bangla pronunciation and uses concatenation based development of Bangla TTS. And they did not mention about the total audio file requirement. In this paper we propose a formal framework based on the basic normalization rules of Bangla grammar. We also minimized the number of audio files to get the output.

III. Bangla Text to Speech Synthesis

A. Text Normalization

Text normalization is the process of transforming text into a single standard form that guarantees to be consistent before operations are performed on it. It is used when converting text to speech, numbers, dates, acronyms, and abbreviations that need to be pronounced differently depending on context [12]. Appropriate normalization can produce accurate speech. In this paper, after splitting the input text, we apply the normalization rules on the intermediate text. Pronunciation of the letters differs according to their position in Bangla. Most of the time, a consonant is used with a vowel and several kinds of pronunciations are generated. Generally there are three kinds of positions can occur of the constants and vowels in a word namely Constant Vowel (CV), Vowel Constant (VC) and Vowel Constant Vowel (VCV). Sometimes a consonant is mixed with one or more than one consonant (CC, CCC etc.) to create a complex sound [18]. The normalization rules are developed based on Bangla pronunciation [13] and on the basis of the position of the constants and vowels. The algorithms are developed based on the rules.

Based on the position of a letter in a word (i.e. front, middle or end of the word), the pronounciation of the corresponding letter becomes different. The pronunciation is also dependent on the previous letter as well. The variants with the normalization rules for 'অ' are described in Table 1. Table2 shows the variants with the normalization rules for ব-ফলা ,ম-ফলা ,য-ফলা ,ল-ফলা and finally normalization rules for the corresponding হ is shown in Table 3.

TABLE I. TEXT NORMALIZATION FOR অ BASED ON DIFFERENT POSITIONS

Rule No.	Positions	Rule Description	Pronunciation	Example
1	Front	If resides in front of word and later consonent is followed by vowel ই /ঈ / উ / ঊ or if exists consonentক্ষ / জ্ঞ or if consonants are combined with য-ফলা / ঋ-কার	অ < ও pronounced	অতীত < ওতীত গভীর < গ্যোভীর কন্যা < কোন্+না
2	Middle	(a) If resides in of word middle and later consonent is followed by vowels ই /ঈ / উ / ঊ / ঋ or if exist consonent ক্ষ / জ্ঞ or if consonants are combined with য-ফলা / ঋ-কার	অ < ও pronounced	কাকলি < কাকোলি অদৃক্ষ < অদ্যোক্+খো
		(b) If resides in middle of word and previous character be vowelঅ/আ/এ/ও or consonent are followed by অ/আ/এ/ও	অ < ও pronounced	চেতন < চেতোন বলদ < বলোদ
3	End	(a) If resides in end of the word	ক=ক+অ < ক pronounced	নাক < নাক মান < মান্
		(b) If resides in end of the word but immediate previous character consists ঐ/ঔ /ং/ঃ /ঋ-কার	অ < ও pronounced	শৈল < শোইলো বিংশ < বিঙশো

TABLE II. TEXT NORMALIZATION FOR ব-ফলা ,ম-ফলা ,য-ফলা ,ল-ফলা IN DIFFERENT POSITIONS

Rule No	Positions	Rule Description	Pronunciation	Example
4	Front	If the first consonant of the is combined with ব-ফলা / ম-ফলা	ব/ম Not pronounced	শ্বাস < শাশ ধ্বনি < ধ্যোনি স্মরণ < শরণ
5	Any	If there exists two consonants combined and followed by ব-ফলা / ম-ফলা /য-ফলা	ব/ম/য Not pronounced	তত্ত্ব < তত্+তো উজ্জ্বল< উজ্+জল যক্ষ্মা < জক্+খা স্বাস্থ্য < শাস+খ্যো
6	Any	If ব-ফলা is specifically combined with consonant গ	ব pronounced normally	দিগ্বিজয়< দিগ্ +বিজয়
7	Middle or End	If ব-ফলা/ম-ফলা/য-ফলা is with other consonents in middle or end of word	Joint letter word repetition	বিশ্বাস < বিশ+শাস ছদ্ম < ছদ্+দো নব্য < নোব্+বো
.8	Any	If ম-ফলা is specifically combined with consonant গ/ঙ/ট/ন/ম/ল	ম pronounced normally	বাগ্মী < বাগ্+মী সম্মান <সম্ +মান
9	Front	If the first consonant of the is combined with য-ফলা	ব্য < ব্যা pronounced	ব্যক্ত < ব্যাক্+তো ব্যাহত < ব্যাহোতো
10	Front	Ifstarting consonents of word are followed byল-ফলা	Normal Pronouciation	গ্লানি < গ্লানি ক্লেশ < ক্লেশ
11	Middle or	If ল-ফলাis with other consonents in middle or end	ল pronounced normally	আপ্লুত < আপ্+লুতো

	End	of word		

TABLE III. TEXT NORMALIZATION FOR হ IN DIFFERENT POSITIONS

Rule No.	Positions	Rule Description	Pronunciation	Example
12	Any	If হ is followed by ঋ-কার/র-ফলা/রেফ	হ < র pronounced	হৃদয় < রিদয়
13	Any	If হ is followed by য-ফলা	হ < য and Joint letter word repetition	উহ্য < উজ্+যো
14	Any	If হ is followed by ল-ফলা	হ্ল < লহ Pronunced	আহ্লাদ < আল+হাদ

B. Audio File Generation

There are 39 consonants and 11 vowels in Bangla language. Vowels can be pronounced independently while consonants cannot be pronounced independently. Normally any consonant with a vowel make one unique syllable. For example ক+আ=কা. But there have some consonants that need not to take help from any vowel to make unique syllable such as ঙ, ঞ etc. Some letters generate same sound such as শ, স, ষ . So we replaced them with a single one. A list of similar letters is shown in table 4.

TABLE IV. REDUCING AUDIO STREAM

Letters	Selected Pronunciation
ই,ঈ	ই
ণ, ন	ন
র, ড়, ঢ়	র
শ, স, ষ	স
ত, ৎ	ত
জ, য	য

After replacement we have 28 independent consonants and 8 (the vowel 'অ' is the exception) vowel hence we create 224 (28*8) audio files for the syllables. An example is shown in table 5 for creating syllables of the consonant ক . It is to be mention that 'অ' does not make any syllable 'অ' is merged with 'ক' (ক+অ=ক). Beside this we need to create diphones i.e. when vowels are placed in front of consonant. We used 28 consonant against 5 vowels to generate 140 (28*5) diphones. Table 6 shows and example of diphone creation for the consonant ক. In summary, we need (9 vowels, 28 consonants, 224 syllables and 140 diphones) 401 audio files to be created.

TABLE V. POSSIBLE COMBINATION OF 'ক'

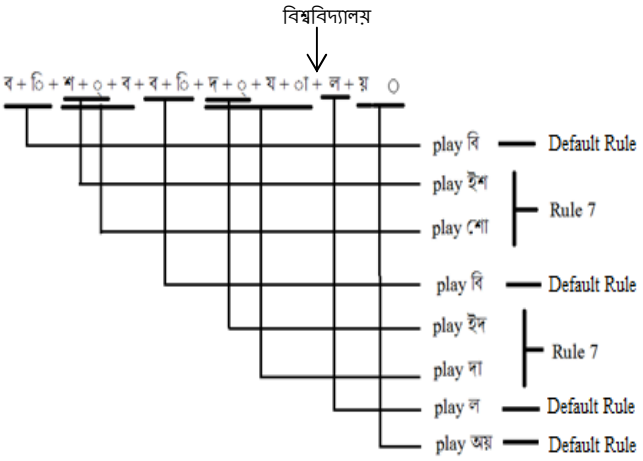
vowel	syllable
া	কা
ি/ী/	কি
ু/ূ/	কু
ৃ	কৃ
ে	কে
ৈ	কৈ
ো	কো
ৌ	কৌ

TABLE IV. POSSIBLE DIPHONE COMBINATION OF 'ক'

vowel	diphone
আ	আক
ই	ইক
উ	উক
এ	এক
ও	ওক

C. Syllable Parser

Each letter in a word is given an index number i. Then we check the (i-1)th letter and (i+1)th letter of the indexed letter according to the normalization rules presented in table 1 , 2 and 3. Suppose the 'ন' letter of "জনতা "word comes to be synthesized. If the index of letter 'ন' is i then 'জ' is i-1 and the letter 'ত' is i+1. After analysis according to rules given above, it is seemed that the pattern of the position of the letter 'ন' is matched with the rule 2(b). So the algorithm picks the লা.mp3 audio file. Figure 1 shows an example of syllable parsing. Table 7 shows some examples for the rule detection procedure that explained in table 2 and 3 with examples.



Syllable parsing of the word **বিশ্ববিদ্যালয়**

IV. EXPERIMENTAL RESULTS

We tested our mythology with the sample test inputs taken from various news articles from news portals and the result is summarized in table 8. The accuracy of the uttered speech is determined by people chosen randomly who heard and understand the paragraph without knowing the input text.

There are some words mentioned in the sample input (see table 9) which causes some problems to the listeners to understand uttered words at first attempt due to duration synchronization and merging between the syllables. For example the word

TABLE VII. RULE DETECTION LOGIC FOR EVERY SYLLABLE IN WORD

SL No.	Condition	Example Word	Reference rule
1	Front অ	জনতা	Rule 1
2	Middle অ	কাকলি ি+ল+ক+া+ক =	Rule 2(a)
3	Endঅ	বাকল ল+ক+া+ব =	Rule 3(a)
4	Front ব-ফলা /ম-ফলা	শ্বাস =শ+্+ব+া+শ	Rule 4
5	ব-ফলাis specifically combined with consonant গ	দিগ্বিজয় =দ+ি+গ+্+ব+ি+জ+য়	Rule 6
6	ম-ফলাis specifically combined with consonant গ/ঙ/ট/ন/ম/ল	বাগ্মী =ব+া+গ+্+ম+ী	Rule 8
7	Front য-ফলা	ব্যক্ত =ব+্+য+ক+্+ত	Rule 9
8	Front ল-ফলা	ক্লেশ =ক+্+ল+ে+শ	Rule 10
9	হis followed by ঋ-কার/র-ফলা/রেফ	হৃদয় =হ+ৃ+দ+য়	Rule 12
10	হis followed by ল-ফলা	আহ্লাদ =আ+হ+্+ল+া+দ	Rule 14

'নম্বর' has been normalized as 'ন+অম্+ব+অর্'. When merging the audio file 'ন.mp3' and 'অম্. mp3' made the duration of pronunciation of 'নম্ব' drow lanigiro eht fo trap 'নম্বর elttil ' of length of duration daviation from original lengthy causing the pronounciation of'নম্বdrow lanigiro eht fo trap '. But after hearing two or three times they were able to determine what has been uttered. The methodology has not been prepared for numerical value written in Bengali. That is why listener did not hear the numerical value written in Bengali. We are working on the system to develop a complete TTS.

TABLE IX. EXPERIMENTAL ANALYSIS

SL. No.	Sample Input	Sentence & Word	Accuracy
1	ওয়ানডের র‍্যাঙ্কিংয়ের সেরা অলরাউন্ডারের স্থানটা বারবার হাতবদল হচ্ছে মোহাম্মদ হাফিজ আর সাকিব আল হাসানের মধ্যে। গত মে মাসে সাকিবকে পেছনে ফেলে শীর্ষে উঠেছিলেন হাফিজ। তবে খুব বেশি দিন দুই নম্বর হয়ে থাকতে হচ্ছে না সাকিবকে। এক মাসের ব্যবধানেই আবার ফিরে পেয়েছেন শীর্ষ স্থানটা। আইসিসির নতুন র‍্যাঙ্কিংয়ে সাকিবের ঝুলিতে জমা হয়েছে ১৮৭ পয়েন্ট। পাঁচ পয়েন্টের ব্যবধানে পিছিয়ে থেকে দ্বিতীয় স্থানে আছেন হাফিজ। সূত্রঃ প্রথম আলো ***তারিখ: ১৮-০৬-২০১৩***	6 sentences & 61 words	47/61 = 77.04%
2	'বাবা দিবস' ঘোষণার বিষয়টি সনোরার মাথায় আসে ১৯১০ সালে। 'মা দিবস' নিয়ে সেই বছর গির্জায় ভালো ভালো কথা শুনছিলেন তিনি। তখন মনে হয়েছিল শুধু মা নয়, বাবা নিয়েও এরকম একটি দিন থাকলে ভালো হয়। এই চিন্তা থেকেই সনোরা গির্জার যাজককে ৬ জুন তার বাবার জন্মদিনটি 'বাবা দিবস' হিসেবে উদযাপনের জন্য প্রস্তাব দেন। যদিও যাজক প্রতিবছর জুন মাসের তৃতীয় রবিবার বাবা দিবস পালনের ঘোষণা দেন। সূত্রঃ bdnews24.com ***তারিখ: ১২-০৪-২০১৩***	5sentences & 60 words	49/60 = 81.66%
3	বৃষ্টিতে কতবার ধুয়ে-মুছে গেছে দক্ষিণ আফ্রিকার স্বপ্ন! ১৯৯২ বিশ্বকাপে ১ বলে ২১ রানের ট্র্যাজেডি। ২০০৩-এ বৃষ্টি আইন বুঝতে না পেরে বিদায়। এর আগে বিদায় নিতে হয়েছিল ১৯৯৯ বিশ্বকাপ সেমিফাইনালে অস্ট্রেলিয়ার সঙ্গে টাই করে। অভিশপ্ত সেই বৃষ্টি আর টাই গলা ধরাধরি করে এবার এলো আশীর্বাদ হয়ে। ওয়েস্ট ইন্ডিজের সঙ্গে টাই করে সেমিফাইনালের টিকিট পেল দক্ষিণ আফ্রিকা। দুই দলের পয়েন্ট সমান হলেও শ্রেয়তর রানরেটে শেষ চারে ভারতের সঙ্গী হয়েছে প্রোটিয়ারাই। সূত্রঃ কালের কণ্ঠ ***তারিখ: ১২-০৪ -২০১৩***	7sentences & 65 words	50/ 65= 76.92%

TABLE X. EXPERIMENTAL ANALYSIS

Sample1	র‍্যাঙ্কিংয়ে, অলরাউন্ডারের,স্থানটা, মোহাম্মদ, নম্বর, পয়েন্ট, শীর্ষ
Sample2	গির্জায়, চিন্তা, ১৯১০, ৬, জন্মদিনটি, তৃতীয়, প্রস্তাব
Sample3	বৃষ্টি, ট্র্যাজেডি, ১৯৯২, ১৯৯৯,১, ২০০৩, অস্ট্রেলিয়া, ওয়েস্ট, ইন্ডিজের, শ্রেয়, সঙ্গী, প্রোটিয়ারা, অভিশপ্ত,আশীর্বাদ

V. CONCLUSION

We explain a formal framework for building text processing components of TTS systems in Bangla language. In the proposed technique, text normalization and synthesis is based on the Bangla syllable pronunciation rules. Beside this the normalization rules presented in this paper have some

exceptions. For example, to detect Noun and Adjective words namely গাছ) Noun) andভালboth words should follow (Adjective) the rule3(a) But they donot follw the rule .3)a (and their pronounciation is different .We believe the proposed framework can be useful for Bangla TTS development to detect the Bangla words with minimum audio file requirement.

REFERENCES

[1] W. Ainsworth , "A system for converting English text into speech", IEEE Transactions on Audio and Electroacoustics, 21:3, pp. 288 – 290, 1973.

[2] T. J. IBM. Watson Res. Center, Yorktown Heights, NY Bakis, R. ; Eide, E.M. ; Fernandez, R. ; Hamza, W. ; Picheny, M.A., The IBM expressive text-to-speech synthesis system for American English, IEEE Transactions on Speech, and Language Processing, 14:4, pp.1099 – 11082006 .

[3] P. Taylor, "Text-to-Speech Synthesis", Cambridge University Press, 2009.

[4] T. Dutoit, "An Introduction to Text-To-Speech Synthesis", Kluwer Academic Publishers, 1997.

[5] K. M. Azharul Hasan, Md Sajidul Islam, G M Mashrur-E-Elahi, Mohammad Navid Izhar, "Sentiment Recognition from Bangla Text", Technical Challenges and Design Issues in Bangla Language Processing, IGI Global, 2013.

[6] E. Moulines, Francis Charpentier, "Pitch synchronous waveform processing techniques for text-to-speech synthesis using diphones", Speech Communication, 9:5, pp. 453–467, 1990.

[7] E. Moulines, CNET LAA/TSS/RCP, Lannion, France Emerard, F. ; Larreur, D. ; Le Saint Milon, J.L. ; Le Faucheur, L. ; Marty, F. ; Charpentier, F. ; Sorin, C., A real-time French text-to-speech system generating high-quality synthetic speech, Acoustics, Speech, and Signal Processing, 1990. ICASSP-90., 1990 International Conference on 309 - 312 vol.1 3-6 Apr 1990.

[8] F Yvon, P Boula de Mareüil, C d Alessandro, V Aubergé, M Bagein, G Bailly, F Béchet, S Foukia, J -F Goldman, E Keller, D. O Shaughnessy, V Pagel, F Sannier, J Véronis, B Zellner, Objective evaluation of grapheme to phoneme conversion for text-to-speech synthesis in French, Computer Speech & Language, Volume 12, Issue 4, October 1998, Pages 393–410.

[9] L. Lee ; Nat. Taiwan Univ., Taipei, Taiwan ; Chiu-yu Tseng ; Ming Ouh-Young, The synthesis rules in a Chinese text-to-speech system, Acoustics, Speech and Signal Processing, IEEE Transactions on (Volume:37 , Issue: 9), 1309 – 1320, 1989.

[10] M. Rokibul Alam Kotwal, F. Hassan, M. Nurul Huda, "Speech Feature Evaluation for Bangla Automatic Speech Recognition", Text In: Technical Challenges and Design Issues in Bangla Language Processing IGI Global, 2013.

[11] Frances Alias, Xavier Servillano, Joan Claudi socoro and Xavier Gonzalvo "Towards High-Quality Next Generation Text-to-Speech Synthesis: A multi domain Approach by Automatic Domain Classification", IEEE Transactions on Audio, Speech and Lnguage Processing, Vol 16,No.7, september 2008.

[12] Sproat, R.; Black, A.; Chen, S.; Kumar, S.; Ostendorfk, M.; Richards, C. "Normalization of non-standard words." Computer Speech and Language, 15, pp.287–333, 2001.

[13] M. Islam, "আবৃত্তি নির্মিতি", Precision publication, 2012.

[14] Md. Hanif Seddiqui, Muhammad. Anwarul Azim, Mohammad Shahidur Rahman, M. Zafar Iqbal "Algorithmic Approach to Synthesize Voice from Bangla Text", 5th ICCIT, pp. 233-236, 2002.

[15] N. P. Narendra, K. Sreenivasa Rao, Krishnendu Ghosh, Ramu Reddy Vempada, Sudhamay Maity, "Development of syllable-based text to speech synthesis system in Bengali", Int J. Speech Technology 14, pp.167–181, 2011.

[16] M. Masud Rashid , Md. Akter Hussain, M. Shahidur Rahman, "Text Normalization and Diphone Preparation for Bangla Speech Synthesis", Journal of Multimedia, 5:6, 2010.

[17] S. Mukherjee, Shyamal Kumar Das Mandal," A Bengali Speech Synthesizer on Android OS", Proceedings of the 1st Workshop on Speech and Multimodal Interaction in Assistive Environments, pp. 43–46, 2012.

[18] F. Alam, S.M. Murtoza Habib, Mumit Khan, "Bangla Text to Speech using Festival", Conference on Human Language Technology for Development, pp. 154-161, 2011.

A Wideband 6.9-17.7 GHz CMOS Low-Noise Amplifier with a Gain Flattening Technique

Toahera Abdullah[1*], Apratim Roy[2], Juhi Johana Bashir[1], Tasnuva Noor[1]
[1] Dept. of EECE, Military Institute of Science and Technology, Dhaka 1216, Bangladesh
[2]Dept. of EEE, Bangladesh University of Engineering and Technology, Dhaka 1000, Bangladesh
Email: *toaherapeu@gmail.com

Abstract— **A flat-gain design of an ultra wide-band CMOS amplifier is proposed and simulated in 90 nm CMOS process. Inductive degeneration is applied to reduce noise figure without significantly raising the architecture's power requirement. Additionally, a resistive shunt feedback technique is applied with an RL peaking load to flatten the gain throughout design band. This topology allows the amplifier to have a very large bandwidth of 6.9-17.7 GHz with a low minimum noise figure of 2.26 dB (at 10 GHz). NF also remains below 5.2 dB across the 10.8 GHz bandwidth. The design exhibits gain which peaks to 11.1 dB and has a low power demand of 15.7 mW (from a 1.2 V supply). Comparing the circuit performance to previously published amplifiers shows that it achieves reduction in NF and power dissipation while maintaining a flatter gain in X-K band.**

***Keywords*—Wideband; CMOS amplifier; Flat-gain; Resistive feedback.**

I. INTRODUCTION

The increasing popularity of large bandwidth applications with high spectral efficiency has opened the door for well-funded research on novel ultra wide-band (UWB) techniques [1]. Though these efforts have often been limited to frequencies lower than 10.6 GHz, the use of wireless networks in day to day gadgets like radios, cell phones, and WLAN [2] has led to increased focus on UWB circuits for frequencies outside the popular 3.1-10.6 GHz range [3]. Since this radio technology is able to ensure low power demand for a wide spectrum and has the capability of transmitting at high data rates (with low cost) [4], wideband transmission for higher frequencies such as K_u and K bands has become vital for wireless communication standards.

One of the challenges of designing a high frequency CMOS receiver is to configure the low-noise amplifier (LNA) in the front end without unduly affecting overall noise figure. While designing amplifiers for microwave frequencies, optimization becomes mandatory as the circuit needs to provide high gain, low noise figure, good input-output matching, and low power consumption simultaneously. In this respect, many trade-offs have to be taken into consideration with regard to satisfying front end specifications and design requirements [5]. For example, a boost of power gain may affect the receiver's noise figure and power consumption and, as a result, drain the battery over a shorter time period. In this paper, the design of a low-power LNA operating between X-K bands is considered which can efficiently sustain >8 dB forward gain over a large bandwidth (6.9-17.7 GHz). A resistive shunt feedback technique is applied to two of the active devices to facilitate flatter gain profiles. In addition, an inductor degenerated topology with a shunt peaking RL load is incorporated. Effects of the gain flattening technique are discussed and simulated performance of the design is provided. The amplifier manages a peak S_{21} of 11.1 dB and a minimum input reflection loss of -15 dB within its range of operation. Its reverse isolation always remains better than -48.2 dB over the 10.8 GHz bandwidth with an optimum NF of 2.26 dB at 10 GHz. Finally, a comparison of the architecture's performance with reported circuits is presented to highlight the advantages of the proposed amplifier for X-K band frequencies.

II. CIRCUIT DESCRIPTION

A. Wideband LNA design

The proposed 90 nm X-K band amplifier including the gain flattening technique is presented in Fig. 1. The circuit incorporates an input matching network (built on an inductor degenerated topology), a cascode stage with shunt resistive feedback, a shunt peaking RL load, and an inductively coupled output stage, all driven by a common 1.2 V supply voltage (V_d). The cascoded stage consists of back to back connected transistors M_a (90 μm/0.1 μm) and M_b (90 μm/0.1 μm) powered by a bias current of 8.12 mA. Reverse isolation is improved with the help of this configuration and a small secondary device (M_b) is chosen to reduce parasitic capacitance. The DC blocking capacitors are C_a and C_b which also work as part of the matching networks.

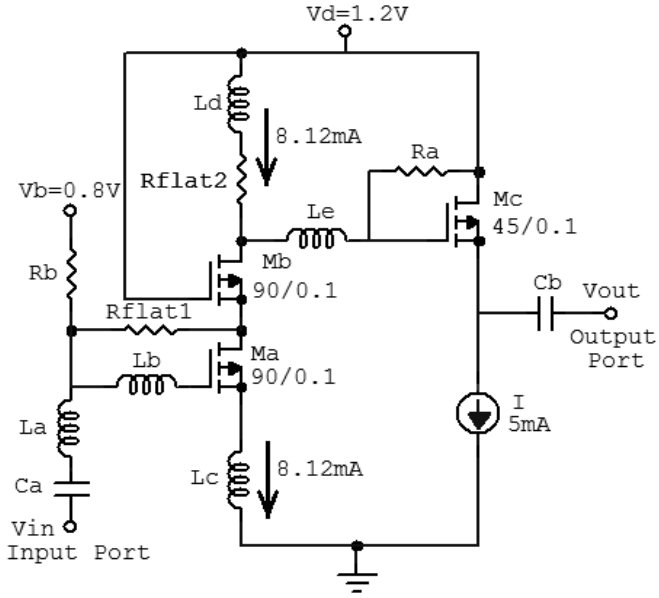


Fig. 1. The proposed flat gain 7-18 GHz low-noise CMOS amplifier.

Resistor R_{flat2} is included with inductor L_d to form the shunt peaking load which also contributes to the gain flattening mechanism described in the following subsection. The output stage is designed with a bias current of 5 mA for the device M_c (45 µm/0.1 µm). For measurement purposes, this output stage will be able to drive an external 50 ohm load [6] by interfacing it with the core amplifier's output impedance.

B. Inductor Degenerated Matching Network

Impedance matching is a crucial design issue for receiver circuits operating at high frequencies. If the impedances of source and input stages are not sufficiently matched, the driving signal is partially reflected back down the line to its source of excitation. Therefore, a wideband input matching network is essential for a front end designed for large bandwidth applications. The components of the inductor degenerated first-stage matching network are inductors L_a, L_b, L_c, capacitor C_a, and transistor M_a, as shown in Fig. 2. Source degeneration and parasitic gate/source elements in this CMOS LNA allow the amplifier to achieve good input matching and optimum noise figure at a resonant frequency. The wideband matching circuit matches input impedance to 50 ohm at the resonant frequency f_1 [see equation (1)] to interface it with an RF source. The series circuit of L_a and C_a provides a second resonated frequency f_2 as obtained from equation (2). The small signal hybrid-π model of the matching network is shown in Fig. 3 for the proposed architecture [7].

$$f_1 = \frac{1}{2\pi\sqrt{L_c Cgs_{Ma}}} \tag{1}$$

$$f_2 = \frac{1}{2\pi\sqrt{L_a C_a}} \tag{2}$$

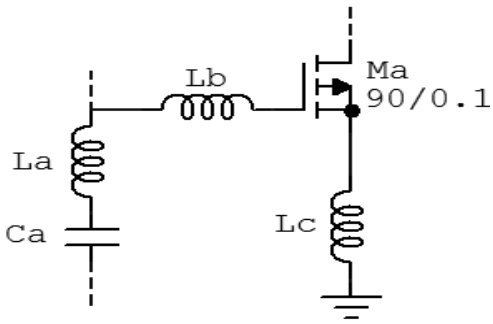


Fig. 2. The driving transistor with inductor degeneration.

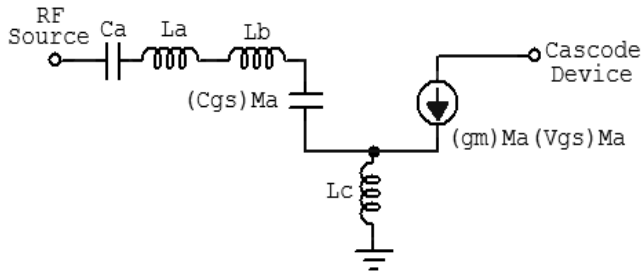


Fig. 3. Small signal hybrid-π model of input network.

C. Resistive Shunt Feedback for Gain Flattening

Though inductor degeneration can be sufficient for optimizing a narrow band design, it can be quite difficult to achieve adequate power and noise matching for a wideband circuit with only source degeneration. Therefore, shunt resistive feedback is introduced to the topology through a feedback connection between transistor terminals in the form of resistors R_{flat1} and R_a. These resistances can control the extent of flatness of the proposed amplifier's bandwidth. To ensure that forward gain remains sufficiently high while bandwidth is regulated, a secondary element R_{flat2} is included with the load inductor L_d. The resistance R_a also provides gate biasing for the output transistor M_c. Appropriate values for R_{flat1} and R_a lead to sufficient matching and low noise figure while maintaining a flat broadband gain performance. The effect of resistive feedback (provided by R_{flat1}) on forward gain is shown in Fig. 4 for a number of combinations. It shows that as R_{flat1} is increased from 700 to 3200 ohm, amplifier bandwidth is enhanced from 10.81 to 11.53 GHz and peak gain is regulated between 10 and 11.5 dB. Fig. 5 presents the regulation of bandwidth which is achievable with the resistive element in the shunt peaking load (R_{flat2}). With a much smaller range of resistance, peak S_{21} can be boosted to values between 10.7 and 12.2 dB while bandwidth is varied within 10.52-10.71 GHz. With the help of these curves, optimal values are chosen for circuit elements to achieve the best possible gain at desired frequencies.

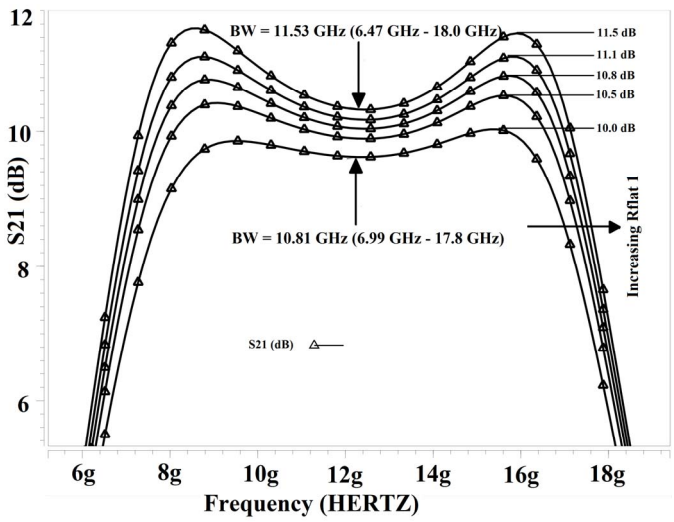


Fig. 4. Regulation of bandwidth achieved by the feedback element R_{flat1}.

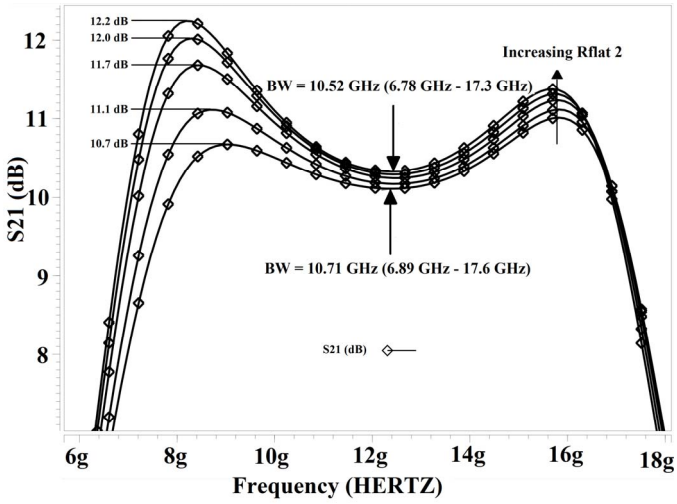


Fig. 5. Effect of resistive element in shunt peaking load (R_{flat2}) on gain.

III. SIMULATION RESULTS

Simulation results of the proposed wideband amplifier obtained using a 90 nm IBM CMOS process are illustrated in this section. As presented in Fig. 6, the optimized amplifier achieves a forward gain (S_{21}) of >8.1 dB with a wide bandwidth of 6.9-17.7 GHz. Fig. 7 shows input (S_{11}) and output (S_{22}) return losses of the architecture. S_{11} stays below -2.6 dB and S_{22} is lower than -0.87 dB over the whole range of operation. Noise figure (NF) is presented in Fig. 8 and it remains between 2.6 and 5.2 dB throughout the operating bandwidth (having a minimum at 10 GHz). Three types of power gain for the design, namely available power gain (G_A), average power gain (G_P), and transducer power gain (G_T), are the subject of Fig. 9. It shows that the range of G_A is 9.61-16.7 dB with a bandwidth of 10.8 GHz. After including the effect of matching networks, the range for G_P and G_T settles at 9.31-12.8 dB and 8.11-11.1 dB, respectively. Input and output voltage standing wave ratios ($VSWR_{in}$ and $VSWR_{out}$) reflect the matching performance of the amplifier in Fig. 10.

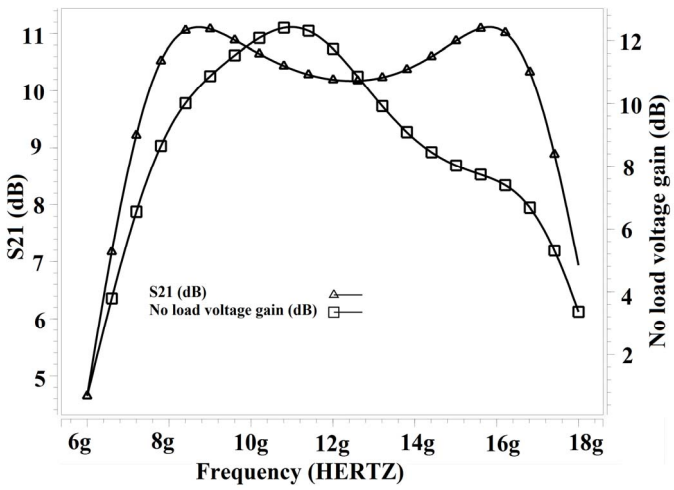


Fig. 6. Optimized voltage gain and forward gain of the amplifier.

The power consumed by the amplifier is 15.7 mW when it is powered from a 1.2 V supply. Table I accumulates performance details of the proposed circuit in terms of microwave parameters and Table II compares its simulated performance against simulated behavior of reported amplifiers with similar noise rating.

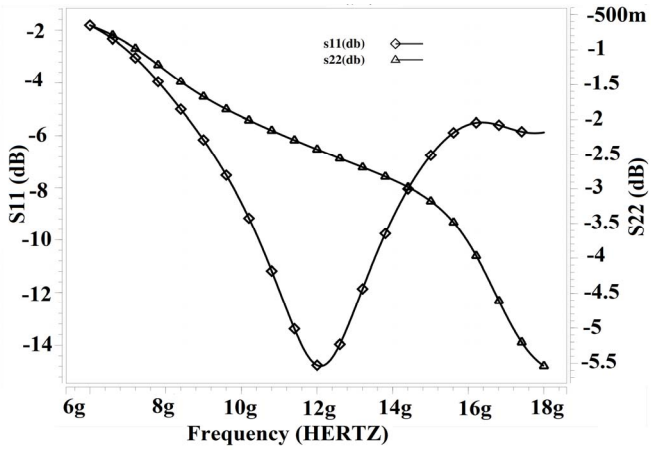


Fig. 7. Input and output return losses indicating port reflection.

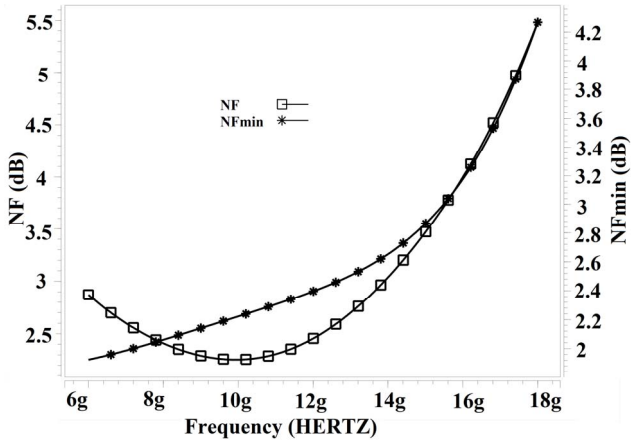


Fig. 8. Amplifier noise figure (NF) and theoretical NF_{min}.

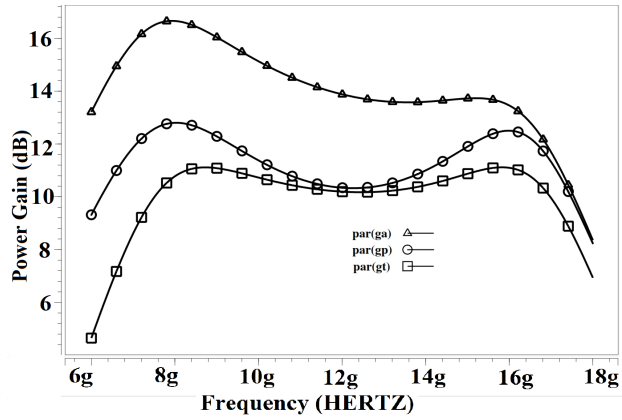


Fig. 9. Simulated power gain of the proposed amplifier.

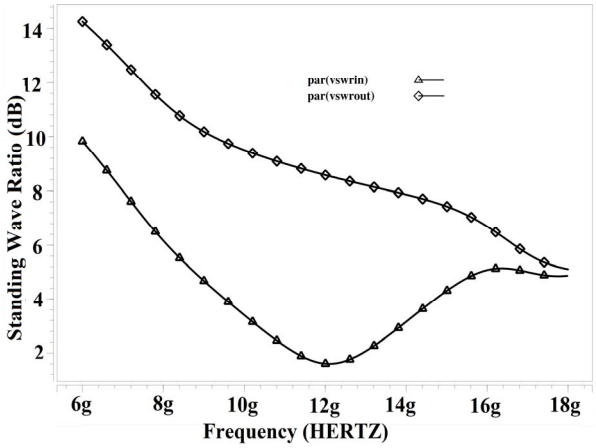


Fig. 10. $VSWR_{in}$ and $VSWR_{out}$ for the circuit.

TABLE I. CHARACTERISTICS OF THE PROPOSED AMPLIFIER.

Parameter (unit)	Value
Supply Voltage (V)	1.2
Voltage Gain (dB)	5.0~12.4
Bandwidth (GHz)	6.9~17.7
Forward Gain (dB)	8.1~11.1
Input Return Loss (dB)	<-2.6
Output Return Loss (dB)	<-0.87
Reverse Isolation (dB)	<-48.2
Power Dissipation (mW)	15.7
Noise Figure (dB)	2.3~5.2

IV. CONCLUSIONS

A CMOS wideband ~15 mW amplifier design is proposed including an inductor degenerated topology and a resistive feedback technique to achieve flat gain. Operating with a wide bandwidth of 6.9~17.7 GHz, the LNA renders a gain of 8.1~11.1 dB over the BW. According to simulation results, input return loss remains below -2.6 dB and output return loss is <-0.87 dB for the concerned frequencies. Minimum noise figure is achieved at 10 GHz and its rating is 2.26 dB. The LNA draws 6.5 mA dc current from a supply of 1.2 V. While comparing simulation results with other published designs, it is noted that this architecture achieves flatter gain at higher frequencies and its noise figure is improved. In conclusion, the proposed gain flattening technique exhibits stable and improved performance for X-K frequency band.

TABLE II. PERFORMANCE COMPARISON WITH OTHER SIMULATED WIDEBAND DESIGNS.

Ref.	**This Work**	[6]	[9]	[10]
Bandwidth (GHz)	**6.9~17.7**	2.3~9.2	3.1~10.6	3.1~10.2
Gain (dB)	**8.1~11.1**	10	7.5~9.7	20
Min. NF (dB)	**2.26**	4	4.5	3.5
P. Diss. (mW)	**15.7**	9	29	24
Supply (V)	**1.2**	1.8	1.8	1.8
Topology	**Gain Flattened**	Inductive Degeneration	Inductive Series	Resistive Feedback
Technology	**90 nm CMOS**	180 nm CMOS	180 nm CMOS	180 nm CMOS

ACKNOWLEDGMENT

The authors would like to thank the authorities of Military Institute of Science and Technology and Bangladesh University of Engineering and Technology for their support and technical assistance.

REFERENCES

[1] B. Allen, M. Ghavami, A. Armogida, and H. Aghvami, "UWB Technology," Commun. Eng., Vol. 1, No. 5, pp. 14–17, Oct. 2003.

[2] Sana Ullah, Murad Ali, Md. Asdaque Hussain, and Kyung Sup Kwak, "Applications of UWB Technology," Publication of Inha University, South Korea, July 2010.

[3] FCC, Rep. and Order, "Revision of Part 15 of the Commission's Rules Regarding Ultra-Wideband Transmission Systems," Feb. 14, 2012.

[4] W. Hirt, "Ultra Wide-band Radio Technology: Overview and Future Research," Comp. Commun., Vol. 26, No. 1, pp. 46–52, Jan. 2003.

[5] Thomas Baker. (2006, Nov). *Successful LNA design involves performance trade-offs* [Online]. Available: http://rfdesign.com /mag/611RFD33.pdf

[6] Andrea Bevilacqua and Ali M. Niknejad, "An Ultra wide-band CMOS Low-Noise Amplifier for 3.1–10.6 GHz Wireless Receivers," IEEE J. Of Solid-State Circuits, Vol. 39, No. 12, pp. 2259-2268, Dec. 2004.

[7] Zhe-Yang Huang, Che-Cheng Huang, Chun-Chieh Chen, Chung-Chih Hung, and Chia-Min Chen, "An Inductor-Coupling Resonated CMOS Low-Noise Amplifier for 3.1-10.6 GHz Ultra Wide-band System," in Proc. of Int. Symp. on Circuits and Syst., pp. 221–224, 2009.

[8] (2011, Oct. 24). *Stability and LNAs* [Online]. Available: http://www.ntms.org/ files/MUD2011/MUD_W5LUA_LNAs_Web.pdf

[9] Chih-Fan Liao and Shen-Iuan Liu, "A broadband noise-canceling CMOS LNA for 3.1-10.6 GHz UWB receiver," in Proc. of the IEEE 2005 Custom Integrated Circuits Conf., pp. 161–164, Sept. 2005.

[10] R. K. Pokharel, A. I. A. Galal, O. Nizhnik, H. Kanaya, and K. Yoshida, "Design of Flat Gain and Low Noise Figure LNA for 3.1-10.2 GHz Band UWB Applications in 0.18 um CMOS Process," in Proc. of IEEJ internation workshop on AVLSI, pp. 161-164, (Istanbul, Turkey), 2008.

A Robust RGB Channel Based Image Steganography Technique using a Secret Key

Anupam Kumar Bairagi, Saikat Mondal, Rameswar Debnath
Computer Science and Engineering
Khulna University
Khulna, Bangladesh
anupam@cseku.ac.bd, saikatmondal@cse.ku.ac.bd, rdebnath@cseku.ac.bd

***Abstract*— The paper proposes a RGB channel based steganographic technique for images imparting better information security. This technique inserts the information into deeper layers of the selected RGB channel and the position is determined depending on the status of channel and value of the secret key. Pixels of the cover image are selected depending on the environment of the channels and hidden information. The ambiguity of pixel, channel and position selection process increases robustness of the steganographic system. The technique is also less vulnerable to unintentional attacks like image manipulation as data hides in the deeper layer of the pixels. The system uses the RGB channels of the stego-image and the secret key to extract the hidden information. The use of secret key gives another way to secure the information from malicious user. The experiment shows that on average 77.00% pixels of the cover image are used for hiding secret information and produces high PSNR value which indicates the capacity and imperceptibility of the technique respectively.**

Keywords*— *Image steganography, RGB image, Stego-image, Secret key, Robustness, Imperceptibility

I. INTRODUCTION

The rise of the internet is the most important factors of information technology and communication. Communication of secret information is a critical factor that continues to create challenges with increasing levels of sophistication. When communication takes place between parties that are located on the same secure network, these challenges can be considered as manageable. However, in the modern age expectations are that one can travel the world and receive secret information at the same time without jeopardizing the confidentiality of information. In these situations where the involved parties are spatially separate, the security of information cannot rely only on the advanced technologies of secure network but some additional security mechanisms should be incorporated.

Since it is unlikely to have a dedicated network line spanning the width of the globe, communication that takes place between remote users often relies on existing public infrastructure, particularly the internet. A public channel such as World Wide Web (www) is convenient to use for remote communication because of its availability. But security of information is the main concern in public channels. In some cases, virtual private networks (VPNs) can be implemented for secured communications by acting as a private tunnel between two parties [2]. However, a VPN still makes use of the internet to establish the network and once an intruder gains access to a VPN tunnel, s/he access to the entire network [3].

So we need a secured and trust-worthy communication way for transmitting important data. To communicate over an insecure channel, cryptography has been developed as a technique for constructing a secure logical channel over an insecure physical channel [1]. Many cryptographic techniques have been developed in order to secure the information. But cryptography suffers from a number of drawbacks- notably the fact that the mere presence of an encrypted message might be cause for suspicion in itself [4], [5]. Another drawback of cryptography is the limitations that have been enforced by certain governments, which is particularly significant when cryptography is to be used by remote users. Many governments have created laws to either limit the strength of cryptographic systems or to prohibit it altogether [6], [7]. Primarily due to law enforcement's fear of not being able to gain intelligence by information interception [7], the use of cryptography and sometimes even the possession of a cryptographic algorithm- are illegal in certain countries [8].

Steganography is a different technique that attempts to protect sensitive information from unauthorized parties. It is the art of hiding information in ways that prevent the detection of hidden message. Steganographic technologies are a very important part of internet security and privacy on open systems. Steganography is a technology that is used to hide secret information in digital media, thus hiding the fact that secret communication is taking place [9]. By hiding secret information in less suspicious digital media, well-known channels, for example e-mail and social networking sites, are avoided, thereby reducing the risk of information being leaked in transit [10]. Steganography provides more confidentiality and information security than cryptography since it conceals the mere existence of secret message rather than only protecting the message contents.

The goal of steganography is secret communication. Therefore, steganography aims to prevent others from thinking that such communication is taking place. Image files are the most common cover medium used for steganography as resolution in most cases is higher than human perception and data can be hidden in noisy bits or pixels of the image.

Enormous research efforts have been invested in the development of digital image steganographic techniques in the recent times. The goal of these is to enhance communication security by inserting secret message into the digital image, modifying nonessential pixels of the image. In ideal scheme, to keep the stego-image from drawing attention from the intruder, should maintain an indiscernible stego-image quality. If there are more similarities between the cover image and the stego-image, it will be harder for an attacker to find out that the stego-image has important secret data hidden inside it. This way, the secret data is more likely to travel from the sender to the receiver safe and sound.

In this proposed technique, RGB channel of the cover image (JPEG) is used for carrying data. It utilizes the secret key, channel environment and secret message to hide information in one of the three channels. It hides secret data into variable position of deeper layer of the channels and changes the LSBs of Blue and Green components accordingly that causes minimum distortions in the stego-image. The system uses secret key and channel environment of the stego-image to extract the hidden information. The use of secret key protects the system from malicious attackers as well.

The rest of the paper is organized as follows: in section 2 and 3 literature review and proposed method are discussed respectively. In section 4 the experimental result and performance analysis is represented. Finally the paper is concluded with section 5.

II. Literature Review

Image steganography has attracted researchers' attention due to high embedding capacity without perceptible impact to the carrier. The simplest approach of hiding data within an image is called LSB substitution. Seven different image steganography methods are discussed in [11] namely Stego1Bit, Stego2Bits, Stego3Bits, Stego4Bits, StegoColourCycle, StegoPRNG and StegoFridrich. These techniques are simple in both embedding and extracting process but LSB is suspicious to the intruder. Stego1Bit method involves utilizing a single LSB of one of the RGB bytes of a 24-bit image for message concealment. As the value is not changed much, it will not considerably alter the visual appearance of the image. Stego2Bits, Stego3Bits and Stego4Bits methods involve utilizing 2, 3 and 4 LSB bits respectively of one of the RGB bytes of a 24-bits image for data embedding. Though these methods increase the data embedding capacity but resulting image is degraded much than Stego1Bit method. StegoColourCycle method involves cycling through the colour values in each of the pixels in which to store the information. StegoPRNG method involves using a pseudo random number generator to choose random pixels in which to hide the data. This will make the message bits more difficult to find and reduce the existence of patterns in the image. StegoFridrich method involves searching for the closest color to the color of the pixel which has the correct parity bit of the RGB values of close colors.

A pixel value differencing method is proposed in [12], where a cover image is partitioned into non-overlapping blocks of two successive pixels. A value of difference is calculated from the value of two pixels in each block. Secret message is embedded into a cover image by replacing the difference values of the two pixel blocks of the cover image with similar ones, in which bits of embedded data are included. Pixel value differencing steganography is vulnerable to histogram based attacks and a modification for enhanced security is proposed in [13].

Two sided, three sided and four sided side match schemes are proposed in [15]. Two sided side match method uses the side information of the upper and left neighboring pixels in order to make estimates. The three sided side match method utilizes upper and left neighboring pixels and one of the other neighboring pixels. Four sided match method uses the upper, left, right and below neighbors. Both [13] and [15] suffer from the fall boundary problem. A solution of this problem is proposed in [17].

Another LSB based image steganography is presented in [18] that utilize the neighboring information to estimate the amount of data to be embedded into an input pixel of cover image and embed a fixed three bits of information in smooth areas and a variable number of bits are embedded into the edge areas. This method uses some pixels of the image to store too many bits of hidden data but other pixels remain unchanged. So, some pixels are distorted roughly but others become unused. They didn't provide any security issue. A steganographic technique based on LSB array presents in [19] where they took all the LSB bits of the different pixels as an array called LSB array, mapped the encrypted message block to this LSB array, embed where the maximum matched. In fact, a number of images are to be experimented; the image in which the percentage of match is maximum can be used.

A steganographic technique based on RGB intensity values of the pixel is presented in [20]. They took one of the channels as indicator channel and used one or both of the remaining to conceal the hidden data. The last two bits of the indicator channel tell whether the data bits are hidden in the other two channels or not. Another two techniques based on RGB are proposed in [21]. In the first, they modified the technique of [20] by changing the indicator of subsequent pixel. The second one is a random number generation based technique, in which a random number determines the number of LSBs, is used to hide the secret data. Another RGB based technique [22] embeds data up to 4 bits in the channels other than the indicator channel based upon the last 4 bits of the indicator channel. A novel RGB channel based steganography technique is proposed in [14] where both the image and the ciphertext are divided into 8 blocks. One ciphertext block is allocated to be embedded in only one image block by a user defined sub-key. Here 4 LSBs are used to store the ciphertext that are fixed and LSBs are more vulnerable to both unintentional and intentional attacks. Two channels are used for storing data.

Another new LSB based image steganographic technique is proposed in [23] where they insert the hidden data in the green or blue channel depending on the secret key. As LSB is suspicious to the attacker and vulnerable to the unintentional attack, thus it is less secured. An improved LSB based image steganography technique for better information security is proposed in [16] for RGB color images. They embed the

encrypted message in non-adjacent and random pixel locations in edges and smooth areas of image. Here, cryptography is adding time complexity of its huge mathematical calculation and also data in LSB is more suspicious.

LSB embedding is sensitive to image manipulation attacks, such as image cropping, resizing and color space conversion or resampling. LSB embedding is also not resistant to statistical attacks, since the embedded information is easy to detect. A fairly simple statistical attack can be used to detect the presence of embedded information by studying the histogram of a stego-image [24]. The process of changing the LSBs of the cover image during LSB embedding, leads to characteristic artifacts in the image histogram that can be used to indentify that steganography has been used [25]. LSB has low robustness against attacks which try to reveal hidden message and against distortion with high average power [26]. Embedding the message bits in the deeper layer and modify other bits else than LSBs can be a solution of this robustness [26]. But this may introduce an alternation complexity.

The purpose of steganography is to keep others from thinking that a secret message even exists within stego files. However, steganalysts may suspect some images and may detect the existence of hidden messages within such images. In order to avoid this, the design of almost all steganographic systems takes into account the characteristic of the human vision system (HVS). Thus, the success of a steganographic system greatly relies on the limitations of the HVS in order to embed data in cover images [15], [10]. As example, if information is embedded in noisy regions and edges of images rather than in smoother regions since the HVS is more sensitive to the degradations in smoother regions.

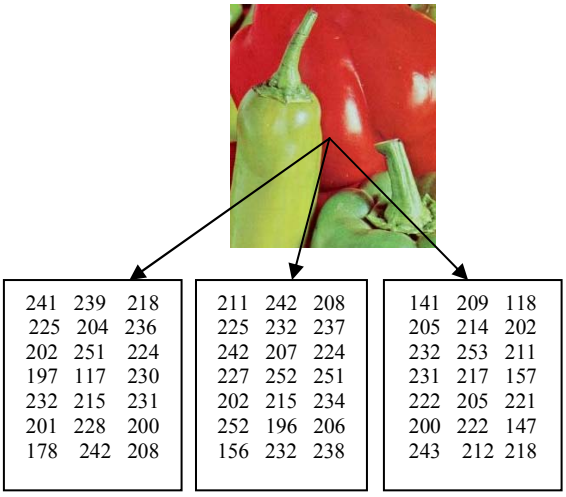


Fig. 1. RGB Matrix representation of cover image

In a computer, images are represented as arrays of values. These values represent the intensities of the three colors R (Red), G (Green) and B (Blue), where a value for each of three colors describes a pixel. Each pixel is a combination of three components (R, G and B) of 8 bits each in a 24-bits color scheme. A cover image can be split into three matrices as shown in Fig. 1. In the proposed method, data bits will be inserted either in the blue component or green component or red component of the image depending on the situation but priority is on the blue component. Blue channel is on top priority because a research was conducted by Hecht, which reveals that the visual perception of intensely blue objects is less distinct than the perception of objects of green and red.

III. Proposed Method

Since substitution techniques usually modify the bits of lower layers in the samples -LSBs, it is easy to reveal the hidden message if the low transparency causes suspicious. Also unintentional attacks like transition distortions could destroy the hidden message if is embedded in the bits of lower layers in the samples -LSBs. Thus embedding the message bits into variable position of the deeper layer of the blue or green or red channel makes it more difficult to discover by the adversary. Moreover, data in deeper layers are less vulnerable to unintentional attacks. Here not all the pixels are used for substitution. They are selected by the situation of the host image pixel and the position of the substitute bit is determined by the status of the bits of blue or green or red component and the secret key. When both the host bit (h) in the determined position in the channel and message bit (m_i) are equal, the pixel is a candidate for data hiding. But when they are not equal in any of the blue or green or red channel, that pixel is not a candidate for data hiding. Here no alternation is made in the deeper layer and only LSB of blue and green channel are used for channel selection. This ambiguity in selecting samples as well as channels will thus increase complexity of analysis and also increase the robustness of the steganographic system. In case of hiding data, priority is given to blue, green and red channels respectively. Only one bit is embedded into a single pixel. The overview of the proposed image steganography technique is shown in the Fig. 2.

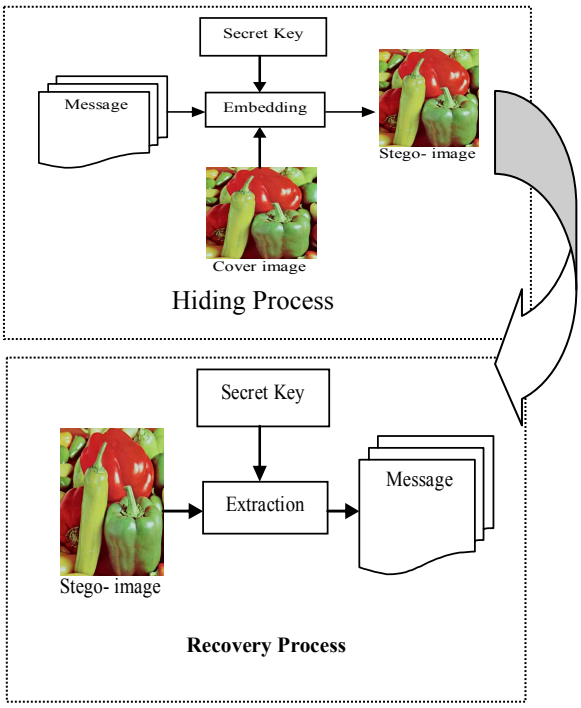


Fig. 2. Proposed image steganography model

A. Data Hiding Technique

To hide the information we have to take a cover image and a message file. This cover image is divided into three matrices (Red, Green and Blue). The message is converted into a 1D array of binary representation and the length of this array is the value of shared secret key (S_k). The number of 1 bit(s) of the sample (except LSB) of the blue, green and red channel is counted and three possible substitution positions in the sample are calculated from the layer 3 to 8 in the three different channels. Now firstly, the message bit (m_i) is checked with the bit of the calculated position in B channel and if they are equal then change LSB of blue to 1 and LSB of green to 0 to indicate B as data channel. Secondly, the message bit (m_i) is checked with the bit of the calculated position in G channel and if they are equal then change LSB of blue to 0 and LSB of green to 1 to indicate G as data channel. Thirdly, the message bit (m_i) is checked with the bit of the calculated position in R channel and if they are equal then change LSB of blue and green to 0 to indicate R as data channel. Here only one channel is used to store the data and priority is given blue, green and red respectively. If the message bit (m_i) is not equal with any of the bit value in the three calculated position in three different channels then change LSB of blue and green to 1 to indicate of no data hiding in the pixel. Thus data is embedding in deeper layer of either blue or green or red channel without changing the value of that position and changing the LSB that is not detectable by human eye. The same process will be continued until the hidden information is finished. The process of message hiding is shown in the Fig. 3.

B. Data Recovery Technique

To recover the hidden information, we have to take the stego-image and the secret key. The stego-image is divided into three matrices (Red, Green and Blue). As LSB of blue and green channel are used as indicator at the time of data hiding, the same is used for identifying the candidate pixel that holds secret bit and also to identify the channel in the pixel. The LSB of blue and green channels (b_1 and g_1) of the sample pixel are X-ORed and if the value is 1 then the secret bit is hidden either in blue or green channel. But if the value is 0 then the secret bit is either in red channel or no data is hidden in the pixel. Depending on the value of b_1 and g_1, the channel (blue, green, red) of hidden bit is selected if the pixel contains any secret data and calculates the exact position in the channel by using the pixel environment and value of the secret key or escape the pixel. The bit of the calculated position in the selected channel is stored in a 1D array. This process will be continued until all the hidden message bits are picked to store in the array. From the 1D array, every 8 bit is separated and converted to character to get the message which was sent by the sender. The process of message extraction from stego-image is shown in the Fig. 4.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

A. Experimental Results

The proposed method has been implemented by using MATLAB R2010a. Five RGB images (JPG) of different sizes, named Lena, Baboon, Pepper, Garden-Home and Player has been used as cover image. The secret message is concealed to the cover image by using proposed hiding algorithm. I have got the stego-images after the embedding process that is shown in Fig. 5 which is closely similar with the cover images. The same key value is used to extract the hidden information from the stego-images by using proposed extraction algorithm into a 1D array that are in the binary form. This binary message is then converted to character by separating 8 bits each to get the actual secret message.

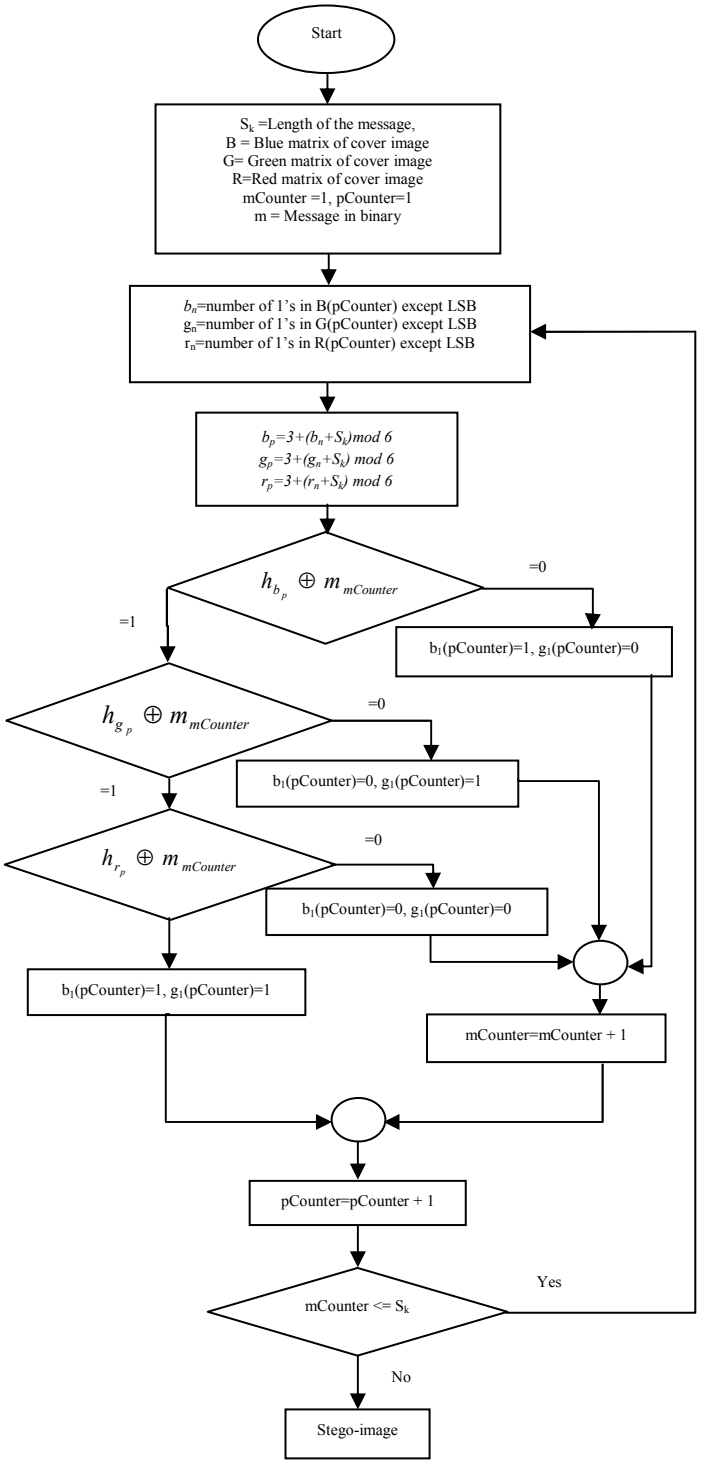


Fig. 3. Data hiding process in the cover image

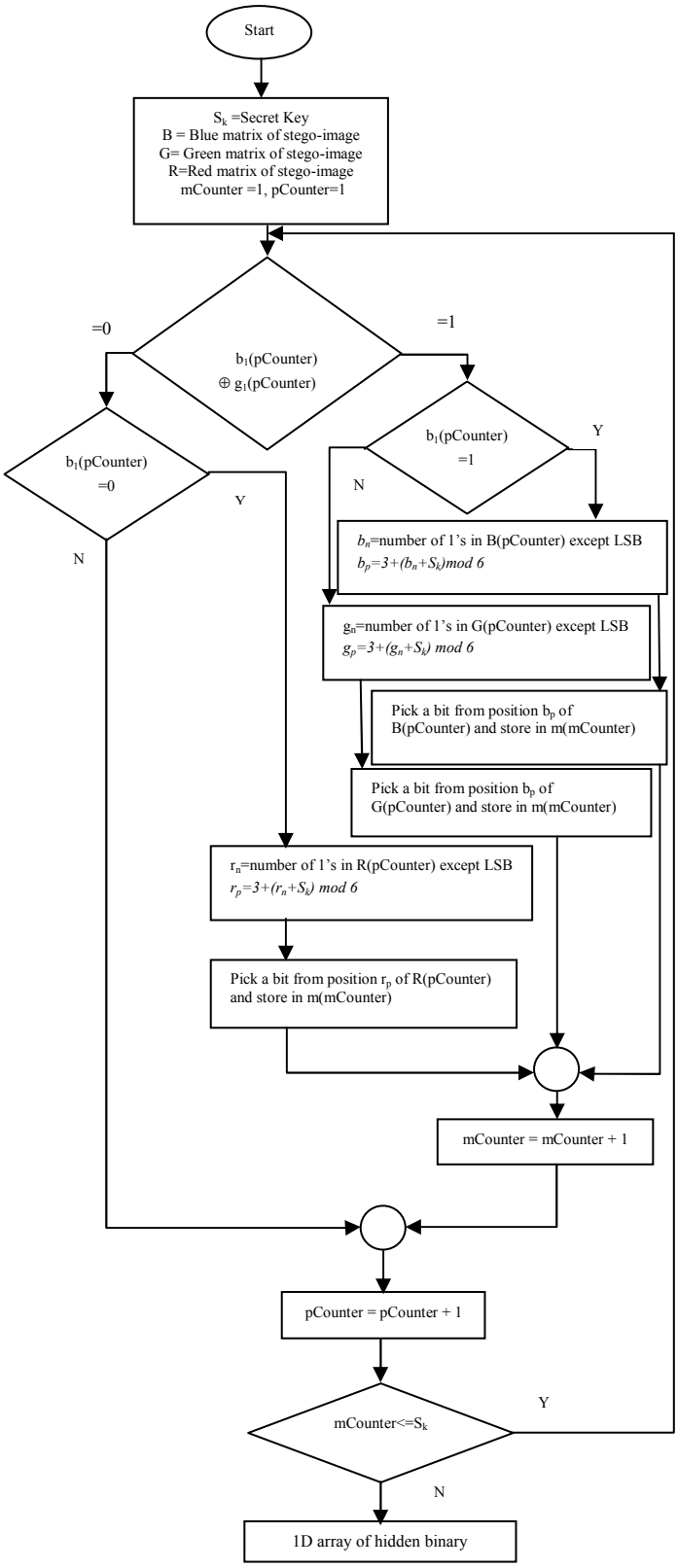


Fig. 4. Message extraction process from stego-image

B. Comarison And Analysis

For measuring the performance of image distortion, the *Peak-Signal-to-Noise Ratio (PSNR)* is calculated. It is defined as:

$$PSNR = 10\log_{10}\left(\frac{C_{\max}^2}{MSE}\right) \quad (1)$$

Where C_{max} holds the maximum possible pixel value in the cover image and *MSE* denotes Mean Square Error which is defined as:

$$MSE = \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}(S_{ij} - C_{ij})^2 \quad (2)$$

Where *i* and *j* are the image co-ordinates, m and n are the dimensions of the image, *S* and *C* are stego-image and cover image respectively. Larger *PSNR* indicates better quality of the image. *PSNR* values falling below 30dB indicate a fairly low quality; a high quality stego-image should strive for 40dB and above [27]. The larger the *PSNR* value the smaller the possibility of visual attack by human eye [18]. The amounts of embedded data and PSNR value depending on "(1)" and "(2)" of the proposed method on different RGB images are shown in the Table 1.

TABLE I. OBSERVED DATA CAPACITY AND PSNR FOR DIFFERENT IMAGES

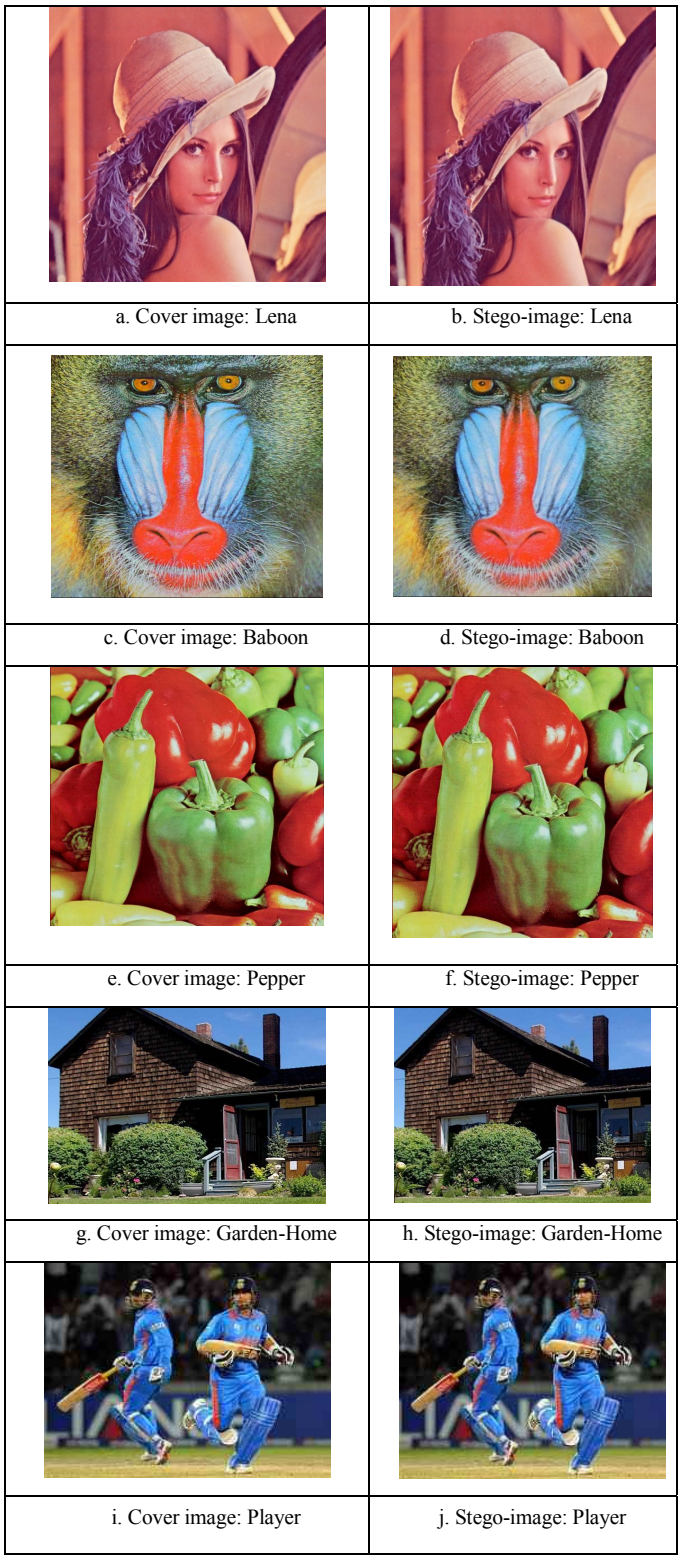
Cover Image	***Image Size (KBs)***	***Data Embedded (bytes)***	***Pixel holds data (%)***	***PSNR (dB)***
Lena	768	15064	85.07	53.7747
Baboon	768	15064	85.66	53.7594
Pepper	768	15064	83.05	53.7995
Garden-Home	880	17264	73.22	52.4846
Player	148	2560	57.94	51.6441
Average			77.00	53.0925

From Table 1, we see that on average approximately 77.00% of the cover image pixel can be used to store the secret information and average PSNR value is 53.0925. A comparison of PSNR values on different images of the proposed system with that of [14] and [16] are shown in the Fig. 6 and Fig. 7. The experimental results show better PSNR values of the proposed technique on different images in comparison with [14] and [16].

The performance of various steganographic methods can be rated by the three parameters: security, capacity and imperceptibility [28]. Steganography may suffer from many active or passive attack, thus a secured steganography technique will have to survive from passive or active attack.

The proposed method is highly secure as it inserts message in variable position in one of the channels and data in deeper layer is less vulnerable to unintentional attack. Pixel, channel and position selection process increase the complexity of steganalysis. From the experiment, we find that high

percentages of pixels are used for information hiding that indicates the data capacity. From Fig. 5, one can observe that there is no visual artifact with the stego-images; they are looking exactly same as their corresponding original cover images.



a. Cover image: Lena | b. Stego-image: Lena

c. Cover image: Baboon | d. Stego-image: Baboon

e. Cover image: Pepper | f. Stego-image: Pepper

g. Cover image: Garden-Home | h. Stego-image: Garden-Home

i. Cover image: Player | j. Stego-image: Player

Fig. 5. Cover image and Stego-image

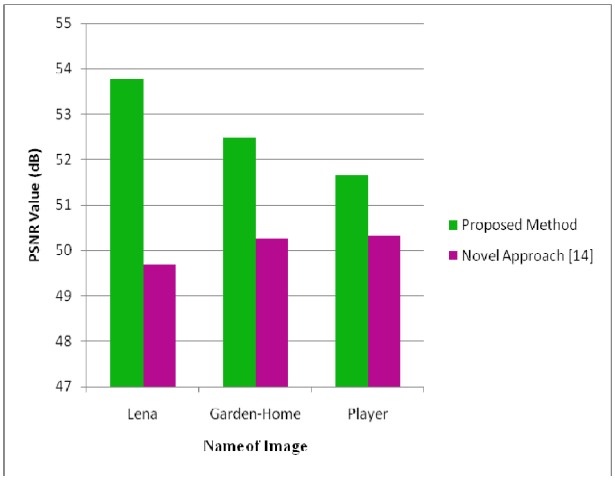


Fig. 6. Comparison of PSNR values of proposed method with [14]

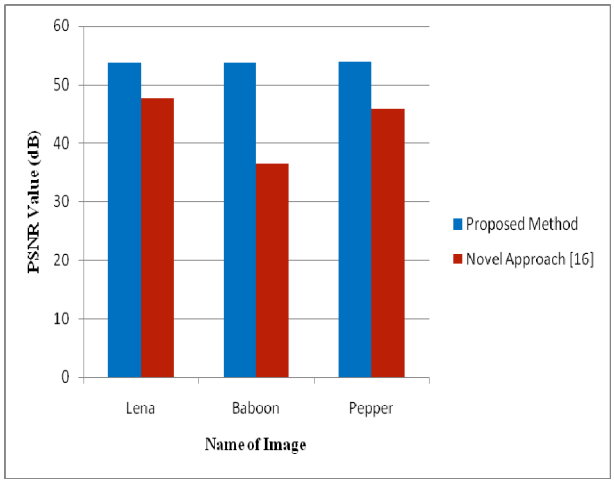


Fig. 7. Comparison of PSNR values of proposed method with [16]

V. CONCLUSION

Steganography aims to protect important data in transmission. In the proposed technique, bits are embedded into variable position of deeper layer where variable position protects data against intentional attacks and deeper layers are less vulnerable to unintentional attacks. Thus the technique has high robustness against intentional and unintentional attack as well. The pixel and channel for substitution is selected by the bit value of the substitution position and message bit. The position of substitution in the channel is determined by the status of selected channel and value of the secret key. The ambiguity in selecting pixels and channel will thus increase complexity of stego-image analysis. The use of secret key gives a way to secure the information from malicious user. The experimental results have shown that the proposed method has good PSNR value which indicates good stego-image quality and high numbers of the pixels of cover image are capable for storing the hidden information. So the proposed technique is a good one to secure the data in transit by increasing the robustness of the steganographic system with imperceptible stego-image quality and good payload.

REFERENCES

[1] D. Gollman,"Computer Security", John Wiley & Sons Publishers, 1999, pp. 201.

[2] W. A. Conklin, G. White,C. Cothren, D. Williams and R. Davis R., " Principles of Computer Security" , McGraw-Hill Technology Education, 2004.

[3] N. Mavrakis, "Vulnerabilities of ISPs: an overall look", IEEE potentials, October/November, 2003, 9-15.

[4] M. Shirali-Shahreza, "Stealth Steganography in SMS", Proceedings of IFIP International Conference on Wireless and Optical Communication Network, 2006, pp. 316-321.

[5] E. Kawaguchi and R. O. Eason, "Principles and Applications of BPCS Steganography", Proceedings of SPIE International Society for Optical Engineering, 1999, pp. 464-473.

[6] J. Krinn,"Introduction to Steganography(2000)", http://rr.sans.org/ covertchannels/steganography.php

[7] F. S. Grodzinsky, K. Miller and M. J. Wolf, "The ethical implications of the messenger's haircut: Steganography in the digital age. In: Himma K. E. (ed.) Internet Security: Hacking, Counter Hacking and Society", Jones and Bartlett Publishers, 2007, pp. 205.

[8] B. Dunbar, "A detailed look at steganographic techniques and the use in an open-system environment", SANS Institute, Information Security Reading Room, pp. 3

[9] T. Jamil, "Steganography: The art of hiding information in plain sight", IEEE Potentials, 1999, 18(1): 10-12

[10] D. Artz, "Digital Steganography: Hiding data within data", IEEE Internet Computing Journal, 2001, 5(3): 75-80.

[11] K. Curran and K. Bailey, "An evaluation of image based steganography methods," International Journal of Digital Evidence, vol. 2, no. 2, pp. 1-40, Fall, 2003.

[12] D. C. Wu and W. H. Tsai, "A steganographic method for images by pixel value differencing", Pattern Recognition Letters, Vol. 24, No. 9-10, pp. 1613-1626, 2003.

[13] X. Zhang and S. Wang, "Vulnerability of pixel value differencing steganography to histogram analysis and modification for enhanced security", Pattern Recognition Letters, Vol, 25, pp. 331-339, 2004.

[14] G. Swain and S. K. Lenka, "A novel approach of RGB channel based image steganography technique", International Arab Journal of e-Technology, Vol. 2, No. 4, June 2012.

[15] C. C. Chang and H. W. Tseng, "A steganographic method for digital images using side match", Pattern Recognition Letters, Vol. 25, No. 12, pp. 1431-1437, 2004.

[16] M. Juneja and P. S. Sandhu, "An improved LSB based steganograhy technique for RGB color images", International Journal of Computer and Communication Engineering, Vol. 2, No. 4, July 2013.

[17] K. J. Kim, K. H. Jung and K. Y. Yoo, "Image steganographic method with variable embedding length", Proceedings of International Symposium of Ubiquitous Computing, pp. 210-213, 2008.

[18] M. Hossain, S. A. Haque and F. Sharmin, "Variable rate steganography in gray scal digital images using neighborhood pixel information", Proceedings of 12th International Conference on Computer and Information Technology (ICCIT 2009), 21-23 December 2009, Dhaka, Bangladesh.

[19] M. Junija and P. S. Sandhu, "Designing of robust stegangraphic technique based on LSB insertion and encryption", Proceedings of International Conference on Advances in Recent Technologies in Communication and Computing, pp. 302-305, 2009.

[20] M. T. Parvez and A. A. Gutub, "RGB based variable-bits image steganography", Proceedings of IEEE Asia Pacific Services Computing Conference, pp. 1322-1327, 2008.

[21] N. Tiwari and M. Shandilya, "Secure RGB image steganography from pixel indicator to triple algorithm- an incremental growth", International Journal of Security and its Applications, Vol. 4, No. 4, pp. 53-62, 2010.

[22] M. Kaur, S. Gupta, P.S. Sandhu and J. Kaur, "A dynamic RGB intensity based steganography scheme", World Academy of Science, Engineering and Technology, Vol. 67, pp. 833-838, 2010.

[23] S. M. M. Karim, M. S. Rahman and M. I. Hossain, "A new approach for LSB based image steganography using secret key", Proceedings of 14th International Conference on Computer and Information Technology (ICCIT 2011), 22-24 December, 2011, Dhaka, Bangladesh.

[24] J. Fridrich, "Steganography in digital media: Principles, algorithms and applications", Cambridge University Press, 2010.

[25] L. Xi, X. Ping and T. Zhang, "Improved LSB matching steganography resisting histogram attacks", Proceedings of the IEEE International Conference on Computer Science and Information Technology, 2010, 203-206.

[26] M. Zamani, A. A. Manaf, R. B. Ahmad, A. M. Zeki and S. Abhullah, "A Genetic-Algorithm-Based Approach for Audio Steganography", World Academy of Science, Engineering and Technology, 54, 2009.

[27] A. Cheddad, J. Condell, K. Curran and P. M. Kevitt, "Digital Image Steganography survey and analysis of current methods", Signal Processing, Vol. 90, pp. 727-752, 2010.

[28] B. Li, J. He, J. Huang and Y. Q. Shi, "A survey on image steganography and steganalysis ", Journal of Information Hiding and Multimedia Signal Processing, Vol. 2, No. 2, pp.142-172, 2011.

An Approach for Enhancing Message Security in Audio Steganography

Ashis Kumar Mandal
Dept. of Computer Engineering
Hajee Md. Danesh Science and Technology University
Dinajpur, Bangladesh

Md. Olioul Islam
Dept. of Telecommunication and Electronic Engineering
Hajee Md. Danesh Science and Technology University
Dinajpur, Bangladesh

Mohammed Kaosar
School of Computing and Mathematics
Charles Sturt University
Bathurst, NSW ,Australia

Md. Delowar Hossain
Dept. of Computer Science and Information Technology
Hajee Md. Danesh Science and Technology University
Dinajpur, Bangladesh

***Abstract*— Concealing a message and ensuring its security is inevitable in data transmission. Among various concepts, one approach is steganography that encodes secret message in indiscernible way. In this paper, we present an audio steganographic technique and propose a novel approach to hide data in the least significant bit (LSB) of the stereo-audio samples with CD-quality. Here, on the basis of stego-key and its parity, message bits are encoded into cover audio samples. In terms of security and imperceptibility, this method is a significant improvement of LSB method for hiding information in audio.**

Index Terms —least significant bit, audio steganography, stego-key, parity bit, message security.

I. INTRODUCTION

Steganography has been used for decades to exchange secret message between two parties in such a manner that, existence of the secret message is concealed from the third party. This is usually done by embedding the secrete message into a cover medium. On the basis of system's requirement, both the secret message and the cover medium can be of any data format, including text, image, audio, and video [1]. It is always important in steganographic technique that the hiding of secret data into cover medium should not make any undesirable change to the cover medium so that the originality of the medium is preserved. Similarly, confidentiality of the message embedded in the cover medium should be strong enough for an imposter to reveal. In audio steganography, the concept is to hide message inside an audio signal. Embedding data would modify the signal which should not be perceptible to the human ear.

In this paper, we propose a novel steganographic system based on LSB technique using uncompressed audio file (wave file) while maintaining high perceptual transparency of resulting audio signals and confidentiality of the message.

Wave files have a master RIFF chunk which includes a WAVE identifier followed by sub-chunks. The data are stored in little-endian byte order. While embedding the secret data, the header part of the wave file (first 44 byte) should be left untouched since change of header may corrupt the audio file [2].

The paper is organized in following orders: section II discusses about some related works in the field of steganography, section III presents our proposed solution, discussion and analysis of the proposed solution appears in section IV and finally, section V concludes the paper.

II. BACKGROUND AND RELATED WORK

Numerous steganographic techniques have been presented for hiding data in audio such as LSB encoding, phase coding, echo hiding and spread spectrum [3].

Human Auditory System (HAS) cannot recognize the phase change in audio signal as easy as recognizing noise in the signal. Phase coding use this fact and encode secret message bits by phase shifting in the phase spectrum of a digital signal [4].

In the case of audio steganography, the spread spectrum method attempts to spread secret information across the audio signal's frequency spectrum as much as possible [5].

In echo hiding, information is embedded in a sound file by introducing an echo into the discrete signal. After the echo has been added, the stego signal retains the same statistical and perceptual characteristics. Data are hidden by manipulating three parameters of the echo signal: the initial amplitude, the offset (delay) and the decay rate so that the echo is not audible [6].

However, the most common, simple and popular approach is LSB insertion method that embeds message in the least-significant bit of cover audio [7]. Initially, sampling technique followed by quantization converts analog audio signal to digital binary sequence. Then LSB of binary sequence of each sample of digitized audio file is replaced with binary equivalent of secret message. As for example, let us consider we want to hide the letter 'a' (ASCII code of 'a' is 97, which is 01100001 in binary) inside 16-bit CD quality sample. First the secret information 'a' and the audio file are converted into bit stream. The least significant column of the audio file is replaced by the bit stream of secrete information 'a'. The resulting file after embedding secret information 'a' is called

stego-file. This technique works well not only for audio but also for image and video steganography, because the changes made to the LSB bit will not be reflected in the final stego signal. Nonetheless, this simple technique does not always ensure high level of security. Therefore, modifications are done to the existing LSB method to improve security and other parameters such as complexity, computational load, SNR (Signal to Noise Ratio), BER (Bit Error Rate), efficiency, etc.

Ensuring message security, Enhanced Audio Steganography (EAS) technique was proposed in [8], where additional layer of encryption and decryption is implemented to enhance robustness and transparency of steganograpy. Here, message in characters is mapped into bits and then encrypted before encoding. In [9] one such approach has been proposed too. This approach deals with the key exchange between the sender and the receiver in such a manner that, the information about the audio sample points, where the data are hidden, can be received by the receiver secretly. Another proposal is Kali's technique [10], in which Fibonacci series associate with encryption algorithm has been used to encrypt the secrete data embedding in audio fie.

Concept of Genetic algorithm has also been applied in audio steganography. In this technique, a smart algorithm embeds the message bits in the deeper layers of samples and alters the other bits to decrease the error. If the alteration is not possible for a sample, it will ignore them. Such a method is proposed in [11].

Embedding data in the higher LSB layers is prone to less attack than those embedded in the lower layers, although it enhances distortion. Nedeljko and Tapio propose such a method in [12]. This method enables the data extending from the fourth to the sixth LSB layers with minimum distortions. Similarly, Mazdak et al. propose a robust steganographic method that embeds the data in multiple, vague and higher LSB layers [13] .
Another research work in [14] uses both parity coding and XORing of LSB. Information is hidden by altering LSBs indirectly considering the parity of samples of cover audio. In the second method, information is hidden by performing XOR operation on LSBs. They also propose encryption along with steganography for providing an additional level of security.

III. Proposed Methodology

In this section we have presented our proposed solution in the sequence of model description, encoding and decoding algorithms, encoding and decoding process explanation.

A. Model Description

Our proposed method uses the stereo-audio file of wave format (16-bit CD quality). At first, it converts the wave file into byte array. Stereo wave file has two channels in its track, the left channel and right channel; secret message will be hidden into both channels. a stego-key or password would determine a channel and exact bit position in a byte where the secret message bit will be placed.

This method introduces a new concept of LSB verification bit which is the LSB of the byte array. The message is mapped into binary bits. The first 50 bits of the array represents the size of the message and the rest of the bits are the secret message. That means
Embedding message = 50 bits for size + secrete message.
This size parameter is important for receiver to reveal the length of a message from stego-audio.

For better understanding of the proposed system, let us consider an example, Bob wants to send a secrete message to Alice using our proposed audio steganographic approach. His embedding message is converted in bit streams and then to message blocks, each of one bit. Similarly, his stego-key is converted into binary, with each block of 4 bits. Now the cover audio is converted in binary sample block where each of it is 16 bits block. Our proposed encoding algorithm takes these three types of input blocks and generates a stego-audio. Bob send this stego-audio to Alice through transmission line. When Alice receives this stego-audio, it is converted into blocks of 16 bits binary streams. She uses the same stego-key, which Bob has used, and it is converted into binary password blocks, with each of 4 bits. Then, our decoding algorithm uses these two types of blocks as input and extracts the message bits. From these embedding message bits original secrete message is generated. And Alice gets the same message that Bob has sent to her. This concept has been illustrated in Fig. 1.

B. Encoding and decoding algorithms

We have considered following notations for our encoding and decoding process:

- M: n-bit long embedding message(50bits+secrete Message) to be transmitted
- M_i: *l*-bit long message block
- B: *k*-bit long password agreed between the sender and the receiver, where *k* mod 4=0
- B_i: 4-bit long password block starting at bit position {(i*4) mod *k*} and ends at bit position {((i*4) +3) mod *k*} in P
- A_c: Cover wave file through which embedding message will be encoded
- S_i^l: 16-bit block of sampled audio bit stream from left channel in A_c. $S_i^l[j]$ indicates j^{th} bit in the block.
- S_i^r: 16-bit block of sampled audio bit stream from right channel in A_c. $S_i^r[j]$ indicates j^{th} bit in the block.
- A_s: Stego wave file where embedding message is kept
- $S_i^{l\prime}$: 16-bit block of sampled audio bit stream from left channel in A_s. $S_i^{l\prime}[j]$ indicates j^{th} bit in the block.

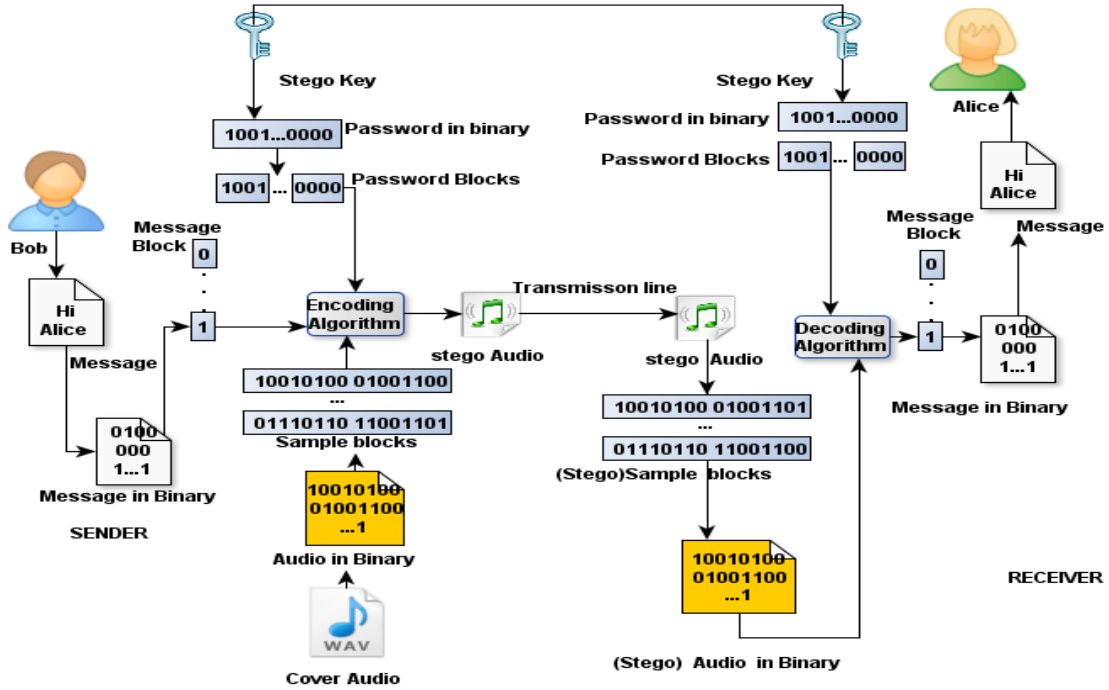


Fig.1. Block diagram of process and flow of information between the sender and the receiver

- $S_i^{r\prime}$:16-bit block of sampled audio bit stream from right channel in A_s. $S_i^{r\prime}[j]$ indicates j^{th} bit in the block.

Following pseudo-code is depicting how message bits are encoded in the sender portion and decoded in receiver portion.

```
GET-PARITY(B_i)
    /*check the parity of a particular
    password block; return 0 if it is even otherwise 1 for odd */
1     return (B_i[0] ⊕B_i[1] ⊕B_i[2] ⊕B_i[3])

FIND-BINARY-DECIMAL(B_i)
    /*convert binary to decimal*/
1     return (B_i[0]*2^0+ B_i[1]*2^1+ B_i[2]*2^2+ B_i[3]*2^3)

ENCODE-OR-DECODE-BIT(Binary x, Binary y, B_i)
1   if FIND-BINARY-DECIMAL(B_i)=0  /*check decimal value
    of B_i */
2       then return 0
3   else
4       return x ⊕ y     /* XOR operation between two
    binary bits*/
```

```
ENCODE-MESSAGE(M, A_c)
    /*Encoding process of message bit*/
1   for i←0 to n-1
2       do P_i ←FIND-BINARY-DECIMAL(B_i)
3         if GET-PARITY (B_i)=1
4         then S_i^l[0] ← ENCODE-OR-DECODE-BIT(S_i^l[P_i], M_i, B_i)
5         else S_i^r[0]← ENCODE-OR-DECODE-BIT (S_i^r[P_i], M_i, B_i)
```

```
DECODE-MESSAGE(M, A_s)
    /*decoding process of message bit*/
1   for i←0 to n-1
2     do P_i← FIND-BINARY-DECIMAL(B_i)
3       if GET-PARITY (B_i)=1
4         then M[i]←ENCODE-OR-DECODE-BIT (S_i^{l\prime}[P_i], S_i^{l\prime}[0], B_i)
5         else M[i]← ENCODE-OR-DECODE-BIT (S_i^{r\prime}[P_i], S_i^{r\prime}[0], B_i)
6   return M
```

C. *Encoding and decoding process explanation*

1) Encoding steps:

In order to hide the message bit M_i into the cover audio sample S_i (=S_i^l or S_i^r) in A_c, parity of block Bi is checked. If the parity is even, it hides the message into S_i^r, otherwise into the S_i^l. Afterward, Bi is converted into decimal number, which is noted by P_i. This decimal value Pi defines a bit position of 0 to 15 in a sample either S_i. The P_i^{th} bit in the sample block S_i represents the parallel message bit M_i. This method does not change the P_i^{th} bit of sample S_i, rather it uses the LSB of sample S_i to verify whether the P_i^{th} bit is matched with M_i or not. If it is matched, the LSB of S_i will be 1, otherwise 0. It is also

important to consider - if decimal value P_i is 0 then LSB of sample S_i is replaced by bit M_i, no matter what the P_i^{th} bit of a sample S_i is. Above mentioned approach changes the LSB of each 16 bit binary number of the A_c, until the end of the embedding message. The steps are shown in Table-I.

TABLE I. CHANGES IN SAMPLE VALUE AFTER EMBEDDING THE MESSAGE WITH CONSIDERING STEGO-KEY

Byte	Sampled Audio stream (16 bits sample) S_i (MSB ..LSB)	Blocked Password (B_i) value in decimal = P_i	Bit at P_i^{th} position of S_i	Message bit M_i	modified Audio stream
Ist and 2nd	10010100 01001100	$1001_2=9_{10}$	0	1	10010100 01001100
3rd And 4th	01010101 11100111	$0011_2=3_{10}$	0	0	01010101 11100111
5th and 6th	00001000 11110001	$1010_2=10_{10}$	0	1	00001000 11110000
7th and 8th	11110000 11000110	$0000_2=0_{10}$	0	0	11110000 11000110

The following figure (Fig.2) depicts the encoding processes of message bits (insertion of 1 and 0) into audio samples.

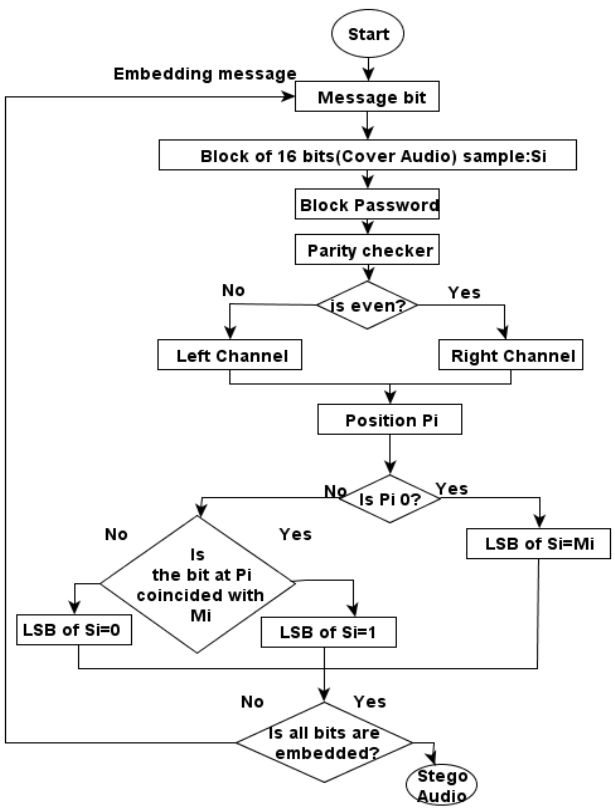


Fig. 2. Flow diagram of embedding the message

2)Decoding steps:

In order to extract the secret message bit M_i from the stego-audio sample S_i' ($=S_i^{l'}$ or $S_i^{r'}$) in A_s, the same password block agreed between the sender and the receiver is used. The password block B_i is then used to select the channel of S_i'. When the parity of B_i is odd, the selection will be the $S_i^{l'}$ to extract message otherwise from $S_i^{r'}$. This block B_i is then converted into decimal value P_i. If P_i becomes 0, LSB of sample S_i' is collected as message bit M_i. Otherwise, the LSB of sample S_i' is collected. If it is 1, from sample S_i' collection of P_i^{th} bit will be message bit M_i; If it is 0,toggling of P_i^{th} bit will be message bit M_i. In this way, this method decodes the M_i .Collection of the first 50 bits and its decimal value indicate size of the message and the iteration will be continue until message size . Table-II shows the extraction of the embedding message.

TABLE II. EXTRACTION OF EMBEDDED MESSAGES FROM STEGO-AUDIO SAMPLE

Byte	Stego-audio stream (S_i') (MSB ..LSB)	Blocked Password (B_i) value in decimal = P_i	LSB of S_i'	Bit at P_i^{th} position of S_i'	Extract Message bit M_i	Extracting message
Ist and 2nd	10010100 01001100	$1001_2=9_{10}$	0	0	1	1010
3rd and 4th	01010101 11100111	$0011_2=3_{10}$	1	0	0	
5th and 6th	00001000 11110000	$1010_2=10_{10}$	0	0	1	
7th and 8th	11110000 11000110	$0000_2= 0_{10}$	0	0	0	

The following figure (Fig.3) depicts the decoding processes of message bits (recovering of 1 and 0) from stego-audio samples.

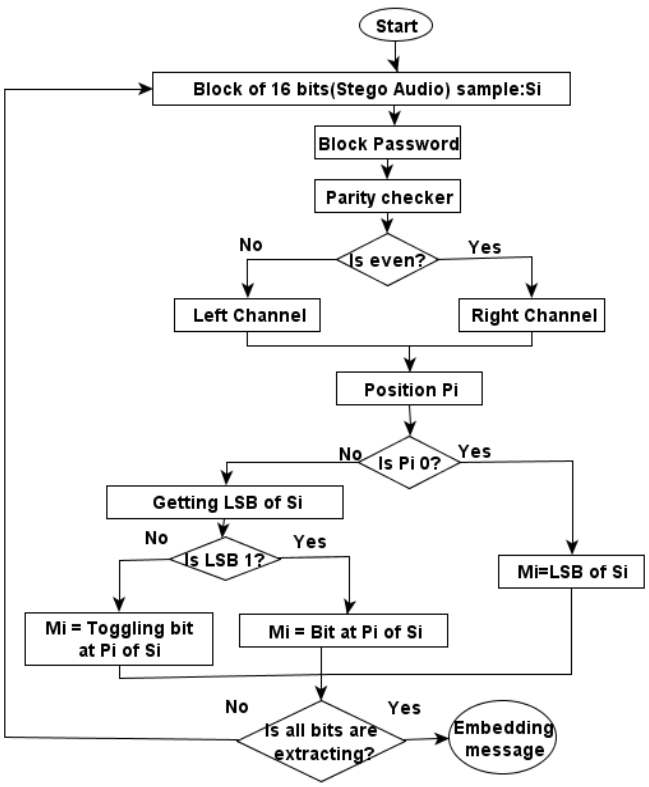


Fig. 3. Flow diagram of extracting the message

IV. Discussion and Analysis

The whole of the system is implemented in C# .Net. To evaluate the performance of the reviewed techniques, ten audio sequences were used as tests signals. Each of the clips (.WAV file) was 44.1 kHz sampled stereo audio files, represented by 16 bits per sample and ranging from 8 to 10 seconds. For enhancing the efficiency of testing, maximum capacity of hidden data are considered. In this system it is 44.1kbps .The approach for testing is the quality evaluation using signal to noise ratio (SNR). As define in [15], SNR for audio steganography in time-domain can be computed using the relation:

$$SNR = 10\log_{10}\frac{\sum_n x^2(n)}{\sum_n [x(n)-y(n)]^2} \qquad (1)$$

where *x(n)* represent sample of cover audio sequence and *y(n)* represent of stego-audio . SNR values for the standard LSB method and the proposed method are given in Fig. 4. It can be seen from this graph that proposed approach has higher SNR value in compared with standard LSB method. Besides, SNR values indicate that after embedding the message, cover audio and stego-audio are approximately alike. It can be clear

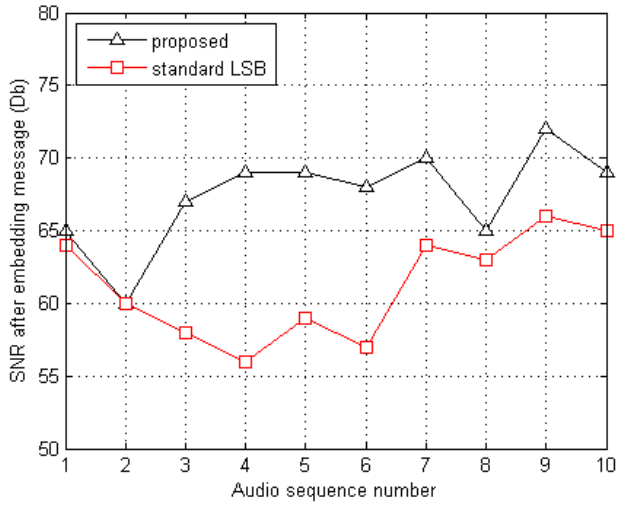


Fig 4. SNR values for 10 test audio sequences after standard and proposed LSB method

from the Fig.5 where we can see the similarity of spectrogram of cover audio signal and stego-audio signal for a certain audio sequence.

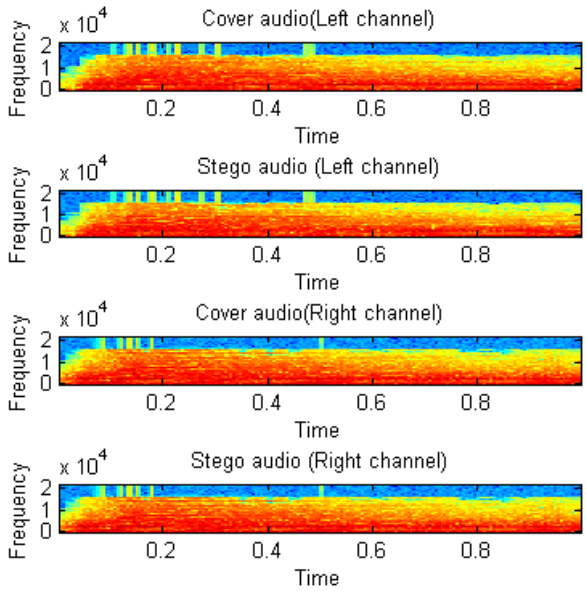


Fig. 5. Spectrogram of cover audio signal and stego-audio signal (Both left and right channel)

The proposed solution has several advantages. First, encoded text message cannot be easily identified because the algorithm alters small number of bits when embedding a large message. And anyhow if a stego-analyzer detect our stego-audio , it is quiet tough to reveal the secret message , for message bits are transformed and placed in the LSB of binary sequence of either left or right channel on the basis of blocked passwords and its parities. In addition, the more the length of the password or stego-key will be, the better the security will be. As the algorithm can use all samples to hide data, a PCM-

coded stereo-audio signal with CD-quality(44100 samples/second; each is encoded 16 bits) of 1 second duration can hide at most around 44100 bits which is of course a big message. Besides, hidden information does not change the size of the cover audio, i.e. stego-audio and clean-audio are alike in terms of size.

Our proposed audio steganographic technique can be applicable for numerous sectors where data security and confidentiality is required. One of the areas is confidential communication where data are pushed into live music or ambient sounds and transmitted over an acoustic channel. This technique can be applied in some applied area, such as simple navigation system, military system and satellite communications for transferring sensitive data. Above and beyond, like watermark, our audio steganographic technique can be used for protecting ownership and controlling copy. For example, to protect a copyright on CD or music, hidden messages into there will allow detection of unauthorized use. Besides, our application can increase data transfer rate as unifying two types of media into one medium.

Although our proposed technique is robust in many cases, the main demerit is that the desire result might not found when stego-audio is altered or affected by noise. Besides, in the case of compressed audio where lossy-compression is occurred, such as in mp3, this process will not provide solution.

V. CONCLUSION

Steganographic approach for secure data transmission has been of a great importance since last few decades. In this paper, we have introduced an ingenious algorithm to conceal message into audio sample. In our approach, we have extended LSB method and have encoded and decoded the message bits in cover audio by stego-key and parity concept. We are optimistic that this would increase the security of the audio steganographic approach in such a way that intruders are less likely to reveal the secrete data.

REFERENCES

[1] Xuping Huang, Ryota Kawashima, Norihisa Segawa, Yoshihiko Abe International Conference on Intelligent, "Information Hiding and Multimedia Signal Processing," The International Journal of Multimedia & Its Applications (IJMA) Vol.3, No.3, August 2011.

[2] https://ccrma.stanford.edu/courses/422/projects/WaveFormat/ (last accessed on 20th February 2013).

[3] Fatiha Djebbar, Beghdad Ayady, Habib Hamamzand Karim Abed-Meraim, "A view on latest audio stegnography," International Conference on Innovations in Information Technology, pp. 409-414, 2011.

[4] M. Nutzinger and J. Wurzer, "A novel phase coding technique for steganography in auditive media," Sixth International Conference on Availability, Reliability and Security (ARES), 2011.

[5] H. Matsuka, "Spread spectrum audio steganography using sub-band phase shifting,"In IEEE International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP06), pp. 36, Pasadena, CA, USA, December 2006.

[6] Erfani, Y. and Siahpoush, S, "Robust audio watermarking using improved TS echo hiding," Digital Signal Processing, vol. 19, pp.809-814, September 2009.

[7] Dr.H.B.Kekre and A.A.Archana, "Information hiding using LSB technique with increased capacity", International Journal of Cryptography and Security, vol. 1, No.2, October 2008.

[8] Zamani, M., Manaf, A., Ahmad, R.B., Jaryani, F., Taherdoost, H., Zeki, A.M., "A secure audio steganography approach", International Conference for Internet Technology and Secured Transactions 2009, Page(s): 1 - 6.

[9] Muhammad Asad, Junaid Gilani, Adnan Khalid "Three Layered Model for Audio Steganography" IEEE 2012.

[10] Kaliappan Gopalan., "Audio steganography using bit modification," IEEE International Conference on Acoustics, Speech, and Signal Processing, Page(s): II -421-4 vol.2,2003.

[11] Mazdak, Z., A.M. Azizah, B.A. Rabiah, M.Z. Akram and A.Shahidan, "A genetic-algorithm-based approach for audio steganography, " World Acad. Sci. 2009

[12] N. Cvejic and T. Seppa□nen,"Increasing robustness of LSB audio steganography by reduced distortion LSB coding," Journal of University Computer Science, vol 11, p56, 2006

[13] Zamani M., Ahmad R.B., Manaf A.B.A., Zeki A.M., "An Approach to Improve the Robustness of Substitution Techniques of Audio Steganography," in Proc. IEEE International Conference on Computer Science and Information Technology, ICCSIT pp: 5-9, 2009

[14] H.B.Kekre, Archana Athawale, Swarnalata Rao, Uttara Athawale, "Information Hiding in Audio Signals," International Journal of Computer Applications,Number 9 - Article 4 ,2010

[15] Y. Hu, P. Loizou, "Evaluation of objective quality measures for speech enhancement," IEEE Transactions on Speech and Audio Processing, 16(1), 229-238, 2008

A novel automatic property weight generator for semantic data integration

Rudra Pratap Deb Nath
Department of
Computer Science and Engineering
University of Chittagong
Chittagong-4331, Bangladesh
Email: rudra@kde.cs.tut.ac.jp

Hanif Seddiqui
Department of
Computer Science and Engineering
Bangabandhu Sheikh Mujibur Rahman
Science and Tech. University, Gopalganj
and University of Chittagong
Chittagong-4331, Bangladesh
Email: hanif@kde.cs.tut.ac.jp

Masaki Aono
Department of
Computer Science and Engineering
Toyohashi University of Technology
Toyohashi,Aichi, Japan
Email: aono@tut.jp

***Abstract*—The dramatic increase of heterogeneous data resources, whether they are semantic knowledge bases or databases, demands for an automatic data integration technique, which is directly affected by the weight of property associated to instances or data. Choosing a suitable metric for generating weight for a property automatically is nevertheless a formidable task. In this study, we analyze different metrics for generating weights and formulate mathematical reasoning to select an appropriate one. Our observation suggests that the weight of a property is highly influenced by the number of instances contain the property, the number of instances contain the distinct value for the property and the total number of instances in a training dataset. The experiments and evaluations illustrate the fact and shows the strength of automatic weight generator for properties in a data integration technique.**

***Keywords*—Instance Matching, Semantic data Integration, Identity Recognition, Record Linkage**

I. Introduction

With the availability of the diversified heterogeneous semantically linked data, integration becomes a key factor to reconcile the data. In semantic web, people, places and things, are connected by means of concepts, properties and their instantiation of domain ontologies. Semantic data means data with semantic i.e. presentation of data with its surrounding information such a formal way so that machine can understand the meaning of the data. Ontology, a explicit, formal, specification of shared conceptualization in a domain, contributes in defining semantic data that is synonymous with ontology instance. However, ontologies in a same domain are often defined differently by different creators influenced by their interest, social behaviors, situation and after all due to their different needs. These connected data are called as instances. Instances are defined in a semantic knowledge base.

A semantic knowledge base contains assertion about instances of two disjoint sets called "concepts", C and "relations" (technically called as *property*), R. The semantic knowledge base is defined as follows:

$$KB = (C, R, I, \imath_C, \imath_R) \tag{1}$$

With the definition of the knowledge base, we find that it consists of two disjoint sets C and R as defined before, a set I whose elements are called instance identifiers (or instance for short), a function $\imath_C : C \rightarrow \Re(I)$ called concept instantiation and a function $\imath_R : R \rightarrow \Re(I^2)$ with $\imath_R(r) \subseteq \imath_C(dom(r)) \times \imath_C(ran(r))$, for all $r \in R$. The function $\imath_R$ is called relation instantiation.

Currently a large number of ontology instances are available in semantic knowledge bases: AllegroGraph[1] [1] contains more than one trillion triples, *a basic building block of Semantic Web formed as* $< subject > < predicate >$ *and* $< object >$, Linked Open Data contains more than fifty billion triples. Due to the proliferation of semantically connected instances, automatic instance reconciliation is getting researchers attention. The problem of instance reconciliation is often called as instance matching problem, also known as semantic data integration.

Ontology instance matching is a relatively new domain for researchers in comparison to record-linkage [2]. Instance matching is an important approach to connect all the islands of instances of semantic world to achieve the interoperability and information integration issues.

In equation 1 for the definition of knowledge base, instances are well defined in terms of properties, which are classified into $DatatypeProperty$ and $ObjectProperty$. An instance may be defined by instantiating properties from a few numbers to hundreds of them. Every property may have different impact on their associated instances. This imposes additional challenges in instance matching techniques.

Most of the instance matching research has been focused on the straight forward instance matching problem with different type of transformations. On the contrary, $HMatch(I)$ [3] tried to address automatically detecting property weight. However, it was only focusing on the distinct value based weight generation, which has a negative impact when there is a small number of available instances containing that property. This paper, in fact, address the necessity of weight of properties, where the authors refer to as featuring properties. Let us give a comprehensive short example. An instance of type "Person" may have property values attached to "hasEmail" and "hasAge" properties. Once two instances has same values

[1] http://www.franz.com/products/allegrograph/

for "hasEmail" property. Both of the instances should be same even if the values against "hasAge" are different as the data might be captured at different year. Therefore, "hasEmail" and "hasAge" have different weight factor. Apparently, "hasGender" may have different weight factor than "hasEmail" and "hasAge" to identify an instance.

In [4] authors propose that properties that have a maximum or an exact cardinality of 1 have a higher impact factor on the matching process. However, it is not true in all cases. For instance, a person has exactly one father i.e. the cardinality of hasFather property is one. However, the father of all siblings is the same one. Therefore, we cant conclude that two persons are same if they have same value for the property hasFather. So far researches in OIM assigns the weight factor to the property in top down approach i.e. by either analyzing the schema of the ontology or manually. However, our effort is to automatically impose the weight by analyzing the information of instances which is more convincing and practical.

We investigated with different factors that affect weight of a property and eventually we found three factors: the uniqueness of the property values, the number of instances a property contains, and the total number of instances in the knowledge base. Obviously, uniqueness of property values has direct relationship with property weight. Combining the three factors we found that property weight is directly proportional to the ratio of the number of distinct values of a property to the number of instances contain that property, which in turn depicts that the number of instances contain a property has a negative effect if the number of distinct values is constant. Moreover, property weight is also directly proportional to the ratio of the number of distinct values of a property to the number of total instances, which in turn give us message that the total number of instances has a negative effect if the number of distinct values is kept constant. Therefore, instead of measuring a straight forward property weight by linear equation may not work properly. Suppose, out of one million instances only ten of them contain birth-date, and unfortunately all of them are unique, in that case, it is not wiser to consider that one million instances must contain unique birth-date. Hence, we propose a metric combining the factors together to generate relatively effective weight factors.

We experimented with the metric of property weight, and we obtained better results comparing with that of considering without weight factor and considering uniqueness only. Our result depicts that our proposed metric for property weight generation has better impact over instance matching technique.

Ontology instance matching is required to compare different individuals with the goal of recognizing the same real-world objects. In particular, the application of instance matching plays an important role in information integration, identity recognition and in ontology population. Ontology schema matching and instance matching work in each other to facilitate in discovering semantic mappings between possibly distributed and heterogeneous semantic data. Identity recognition is a widely used term in database and emerging topic in the semantic web of detecting whether two different resource descriptions refer to the same real-world entity, namely an individual. Ontology population is evolved by acquiring new semantic descriptions of data extracted from heterogeneous data sources. For this ontology population, instance matching plays a crucial role to correctly perform the insertion activity and to discover a set of semantic mappings between new incoming instance and the set of instances already stored in an ontology.Moreover, we can also apply our metric to classical database problem namely record linkage.

The rest of the paper is organized as follows. **Section II** compares our idea with other existing related work to articulate a research gap. The factors that affect the property weight are articulated at **Section III** along with some comprehensive examples. **Section IV** contains the mathematical explanation of our metric and necessity of the different considerable factors. **Section V** includes experiments and evaluation to show the effectiveness of our proposed metric to generate property weight factors to match different instances. Concluded remarks and some future directions of our work is described in **Section VI**.

II. Related Work

The rising demand of sharing knowledge, data and information within same or heterogeneous knowledgebases has recently attained a novel attention on issues related to ontology and instance matching. Until now, many researchers have invested their efforts on ontology instance matching to resolve the interoperability issues across heterogeneous sources. SERIMI [5] matches instances between a source and a target datasets, without prior knowledge of the data, domain or schema of these datasets. However, in the instance matching process, SERIMI does not impose any weights to the properties associated with instances. The weight of each property can be manually specified by a domain expert [3] or it can be automatically determined through statistical analysis[6], [7],[8],[9]. In HMatch 2.0 [3], each property is associated with a weight ranging from 0 to 1 expressing the capability of the property for the goal of univocally identifying the individual in the domain of interest. This weight is defined during the featuring properties identification step of the instance matching process. In BOEMIE [3], property weights are manually defined for the considered domain by taking into account also the results of the extraction process from a corpus of (manually) annotated multimedia resources. Nonetheless, manual definition of weight requires involvement of domain experts and the definition of weight may vary among different domains. To discover semantic equivalence between persons in online profiles or otherwise, an appropriate metric is proposed in [4] for weighting the attributes which were syntactically and/or semantically matched. Factors that have been taken into account by the metric are the total number of attributes that were mapped to the ontology framework, the number of syntactically matched attributes, the number of attributes matched based on the target domain of the specific online account. Besides, ontology-enhanced attribute weights are an added benefit of the ontology framework over other ontologies such as FOAF. Attribute constraints defined in the NCO ontology, such as cardinality and inverse functional properties; enable the assignment of different predefined weights to the attributes. Thus, the properties that have a maximum or an exact cardinality of 1 have a higher impact on the likelihood those two particular profiles are semantically equivalent. Yet, it is not true in all cases. For instance, a person has exactly one father i.e. the cardinality of hasFather property is one. However, father of all siblings is the same one. Thereafter, we cant conclude that two persons are same if they have same

value for the property hasFather. A further refinement of the instance matching process is to take into account the frequency each value occurs [6]. In particular, a pair of matching attribute values will receive a high weight if these values occur with a low frequency within the domain, while they will receive a low weight otherwise.

A different approach to improve instance matching performance is based on the idea that, each instance pair (i1, i2) can be classified as matching or non-matching analyzing only a subset of the corresponding property values of i1 and i2 respectively assume on their most identifying properties. Thus, the matching process can terminate when the knowledge about i1 and i2 is enough to classify them as matching or non-matching instances and further comparisons among their property values would be useless, as their results cannot change the classification choice [10].

In [11] RiMOM has been presented. They used several instance matching benchmark data sets to evaluate their systems namely, A-R-S, T-S-D and IIMB. However, for different datasets, their matching strategy is different.

In [12], J. Huber et al have proposed CODI: Combinatorial Optimization for Data Integration in where they emphasize on object-properties to determine the instances for which the similarity should be computed. Although object-properties have a strong influence in the matching process, involvement of data-properties in the matching process is also necessary. Until now, most of the researches in the ontology instance matching assign the weight factor to the property in top down approach i.e. by either analyzing the schema of the ontology or manually. However, our effort is to automatically impose the weight by analyzing the information of instances which is more convincing, generic and practical.

III. Factors that influence the Property Weight

Instances contain values associating with properties. Some properties like *name*, *date-of-birth*, *homepage* have larger weighting factors than the properties like *height*, *frequency* and so on. However, determining the weight factor automatically is a formidable task. Some properties have great influence on identifying instances, while the other have less influence in a semantic knowledge base. For example, an instance of type "Person" may have property values attached to "hasEmail" and "hasAge" properties. Once two instances has same values for "hasEmail" property. Both of the instances must be same even if the values against "hasAge" are different as the data is captured at different year. Therefore, "hasEmail" and "hasAge" have different weight factor.

A. Influence of Negative Factors

Our basic hypothesis to identify the influence of a property on instance identification is that a property has higher influence if its values do not repeat in a semantic knowledge base like a primary key in a database repository. Alternatively a property has less influence if it repeats in the knowledge base. As many times the property value repeats, it looses its ability to identify an instance. Moreover, if a property represent a small number of instances

If a property value repeats, the weight is penalized by a negative probability factor, np defined as a ration of the number of repetition to the number of instances the property belongs to, i.e.

$$np_1 = \frac{NumberOfValueRepetitionOfProperty\ p, denoted\ by\ |dup|}{NumberOfInstancesContainProperty\ p, denoted\ by\ |i \ni p|} \tag{2}$$

Moreover, property value repetition may also have negative effect over the total number of instance. Primarily, let us consider the fact in the equation below:

$$np_2 = \frac{NumberOfValueRepetitionOfProperty\ p, denoted\ by\ |dup|}{TotalNumberOfInstances,\ denoted\ by\ |I|} \tag{3}$$

B. Our Proposed Property Weight Factors

As described in Subsection III-A, there are two types of negative factors associated with our automatic weight generation for properties of semantic knowledge base, namely np_1 and np_2 and they are defined as primarily as follows:

$$np_1(p) = \frac{|dup|}{|i \ni p|}, \tag{4}$$

$$np_2(p) = \frac{|dup|}{|I|}, \tag{5}$$

where np is the negative factor associated to a property, p, $|dup|$ is the number of value duplication for a property p available throughout the knowledge base and i represents an instance. Moreover, $|i \ni p|$ represents the total number of instances containing the property, p and $|I|$ represents the total number of instances in the knowledge base.

The probability of p would be denoted as $Prob(p) = 1 - np(p)$. We consider the probability as the weight of a property. Therefore, we measure the weight of each property of an ontology schema used in a knowledge base as a joint probability and is stated follows:

$$weight(p) = (1.0 - np_1(p)) * (1.0 - np_2(p)), \tag{6}$$

where np_1 and np_2 are defined above.

IV. Mathematical Explanation

A. Mathematical Derivations

Let us start from the joint probability equation 6 for mathematical derivation and to look insight the nature of the

equation.

$$\begin{aligned} weight(p) &= (1.0 - np_1(p)) * (1.0 - np_2(p)) \\ &= \left(1.0 - \frac{|dup|}{|i \ni p|}\right) * \left(1.0 - \frac{|dup|}{|I|}\right) \\ &= \frac{|i \ni p| - |dup|}{|i \ni p|} * \frac{|I| - |dup|}{|I|} \\ &= \frac{(|i \ni p| - |dup|) * (|I| - |dup|)}{|i \ni p| * |I|} \\ &= \frac{(|i \ni p| - |dup|) * ((|I| - |i \ni p|) + (|i \ni p| - |dup|))}{|i \ni p| * |I|} \\ &= \frac{|distinct| * (|i \not\ni p| + |distinct|)}{|i \ni p| * |I|} \\ &= \frac{|distinct|}{|i \ni p|} * \frac{(|i \not\ni p| + |distinct|)}{|I|} \\ &= \frac{|distinct|}{|i \ni p|} * \left(\frac{(|i \not\ni p|}{|I|} + \frac{|distinct|)}{|I|}\right) \\ &= \frac{|distinct|^2}{|i \ni p| * |I|} + \frac{|distinct| * |i \not\ni p|}{|i \ni p| * |I|} \end{aligned} \tag{7}$$

where $|distinct| = |i \ni p| - |dup|$ is the number of instances that contain distinct values to property p. Although the first term $\frac{|distinct|^2}{|i \ni p| * |I|}$ is quite convincing in equation 7, however in the second term $\frac{|distinct| * |i \not\ni p|}{|i \ni p| * |I|}$, we consider that $|i \not\ni p|$ has a positive contribution to the weight factor, which is a contradiction. Let us consider that there are 1 million instances as human being in a knowledge base and 10 instances of them are containing $date_of_birth$ property values and unfortunately all of them have a distinct value. This obviously does not guaranty that the rest instances will contain distinct values. On the other hand, it is not also guaranteed that most of them are duplicate value. Therefore, we need a factorization parameter, λ before the second term as a multiplier, which is defined as below:

$$\lambda = \begin{cases} \frac{|distinct|}{|i \ni p|} & : \frac{|i \ni p|}{|I|} \geq \delta \\ \frac{4|I| - (|i \not\ni p| + |i \not\ni distinct|)}{4|I|} & : otherwise \end{cases} \tag{8}$$

where δ is the empirical threshold, $|i \not\ni p|$ indicates the number of instances that do not contain property p and $|i \not\ni distinct|$ stands for denoting the number of instances that do not have distinct values for property p.

Therefore, after introducing λ , the equation 7 becomes as:

$$\begin{aligned} weight(p) &= \frac{|distinct|^2}{|i \ni p| * |I|} + \lambda * \frac{|distinct| * |i \not\ni p|}{|i \ni p| * |I|} \\ &= \left(\frac{|distinct|}{|i \ni p|}\right) * \left(\frac{|distinct| + \lambda * |i \not\ni p|}{|I|}\right) \end{aligned} \tag{9}$$

B. Comprehensive Example

Let us consider a number of comprehensive examples to understand the equation 9.

If we consider that a property, p is densely instantiated among instances, i.e. every instance in the knowledge base contains some values of p, then $|i \not\ni p|$ is zero. Hence, the equation 9 becomes:

$$weight(p) = \left(\frac{|distinct|}{|i \ni p|}\right) * \left(\frac{|distinct|}{|I|}\right) \tag{10}$$

Let the total number of instance, $|I|$ and the number of instances having distinct values $|distinct|$ of a property, p be constants. In this case, if the number of instances containing p, denoted as $|i \ni p|$ increases, the possibility of identifying an instance with that property decreases. Because, as the $|distinct|$ remain constants and $|i \ni p|$ increases, therefore, duplicate values increases, which means probability of identifying an instance decreases and hence weight of the property decreases and vice-versa. This scenario depicts the natural effect and is successfully addressed in equation 10.

Let the total number of instance, $|I|$ and the number of instances containing property p denoted by $|i \ni p|$ be constants. In this case, if the number of distinct values, denoted as $|distinct|$ increases, the possibility of identifying an instance with that property increases. Therefore, duplicate values decreases, which means probability of identifying an instance increases and hence weight of the property increases and vice-versa. This scenario also depicts the natural effect and is successfully addressed in equation 10.

Let the number of distinct values, denoted as $|distinct|$ and the number of instances containing property p denoted by $|i \ni p|$ be constants. In this case, if the total number of instance, denoted as $|I|$ increases, the possibility of identifying an instance with that property decreases as the the non-identifiable instances increase, which means probability of identifying an instance decreases and hence weight of the property decreases and vice-versa. This is also addressed in equation 10.

In equation 9, if we consider that a property, p is sparsely instantiated among instances, i.e. some instances in the knowledge base may not contain values of p, then $|i \not\ni p|$ is not zero. Hence, the equation 9 remain as it is.

Let the total number of instance, $|I|$ and the number of instances having distinct values $|distinct|$ of a property, p be constants. In this case, if the number of instances containing p, denoted as $|i \ni p|$ increases, the possibility of identifying an instance with that property decreases. Because, as the $|distinct|$ remain constants and $|i \ni p|$ increases, therefore, duplicate values increases, which means probability of identifying an instance decreases and hence weight of the property decreases and vice-versa.

Let the total number of instance, $|I|$ and the number of instances containing property p denoted by $|i \ni p|$ be constants. In this case, if the number of distinct values, denoted as $|distinct|$ increases, the possibility of identifying an instance with that property increases. Therefore, duplicate values decreases, which means probability of identifying an instance increases and hence weight of the property increases and vice-versa.

Let the number of distinct values, denoted as $|distinct|$ and the number of instances containing property p denoted by $|i \ni p|$ be constants. In this case, if the total number of instance, denoted as $|I|$ increases, $\lambda * |i \not\ni p|$ increases partly, therefore the possibility of identifying an instance with that property decreases as the the non-identifiable instances increase, which means probability of identifying an instance decreases and hence weight of the property decreases and vice-versa.

Therefore, it is now obvious that the term $\lambda * |i \not\ni p|$ does not affect the natural behavior, rather it only reduces the adverse effect of $|i \not\ni p|$.

V. Experiments and Evaluations

A generated benchmark to test the efficiency of an instance matcher is called as ISLab Instance Matching Benchmark(IIMB)[2]. Every year, IIMB knowledge base is undated. The test-bed provides OWL/RDF data about actors, sport persons, and business firms. For 2009 data set, the main directory contains 37 sub-directories and the original ABox and the associated TBox (abox.owl and tbox.owl). The original ABox contains about 222 different instances with a number of associated property values. Each sub-directory contains a modified ABox (abox.owl + tbox.owl) and the corresponding mapping with the instances in the original ABox (refalign.rdf). The benchmark data is divided into four major groups: value transformation (001-010), structural transformation (011-019), logical transformation (020-029) and combination transformation (030-037) [13]. The 2010 edition of IIMB is a collection of OWL ontologies consisting of 29 concepts, 20 object properties, 12 data properties and thousands of individuals divided into 80 test cases. In fact, in IIMB 2010,80 test cases are defined and divided into 4 sets of 20 test cases each. The rst three sets are different implementations of data value, data structure and data semantic transformations, respectively, while the fourth set is obtained by combining together the three kinds of transformations. IIMB 2010 is created by extracting data from Freebase, an open knowledge base that contains information about 11 million real objects including movies, books, TV shows, celebrities, locations, companies and more. The benchmark has been generated in a small version consisting in 363 individuals and in a large version containing 1416 individuals [14]. Here, large version set is considered in evaluation. We perform two independent experiments for our instance mathcer by not considering property weight and considering property weight on the IIMB benchmark data set. The consecutive sections contain the corresponding evaluation respectively.

In the experiment of instance matching, we have conducted evaluations in terms of precision, recall and f-measure as defined below:

- *Precision, P*: It is the ratio of the number of correct discovered aligned pairs to the total number of discovered aligned pairs.

$$P = \frac{|i_r \cap i_{cr}|}{|i_r|}$$

[2]http://islab.dico.unimi.it/iimb/

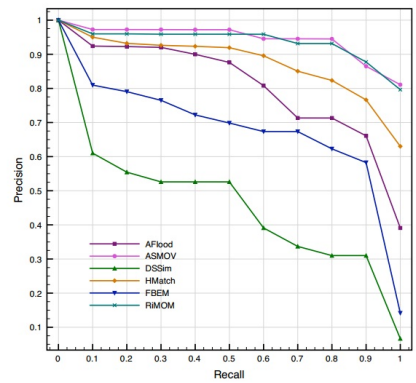


Fig. 1. Instance matching results against IIMB benchmarks. Instance matcher called $aflood$ is our previous algorithm without considering weight factors [13]

- *Recall, R*: It is defined as the ratio of the number of correct discovered aligned pairs to the total number of correct aligned pairs.

$$R = \frac{|i_r \cap i_{cr}|}{|i_c|}$$

- *f-Measure*: It is a measure to combine precision, P and recall, R.

$$f - Measure = \frac{2 * P * R}{(P + R)}$$

where i_r is the total number of retrieved aligned or matched instance pair, i_{cr} is the total number of correctly retrieved aligned or matched instance pair and i_c is the total number of correct aligned or matched instance pair across semantic knowledge bases.

A. Evaluation with IIMB2009 Data set

For the first time, an instance matching track was proposed to the participants in the OAEI-2009. Fig. 1 demonstrates the results of the participants [13] in where *aflood* is our data integrator without considering property weight. Our system was average in performance although it was the fastest in speed. Fig. 2 illustrates the improvement of our system in recall-precision graph after consideration of the proposed automatic weight factor. Here, DI(PW+) stands for Data Integrator with Property Weight and DI(PW-) is used for Data Integrator without Property Weight. Table I shows the result of different transformations when automatic property weight is not considered.

Datasets	Transformation	Prec.	Rec.	F-Measure
001-010	Value transformations	1.00	0.99	0.995
011-019	Structural transformations	0.90	0.72	0.800
020-029	Logical transformations	0.85	1.00	0.919
030-037	Several combinations of the previous transformations	0.94	0.75	0.834

TABLE I. Instance matching results against IIMB benchmarks without weight factors

The recall, precision and f-measure of OAEI2009 dataset by considering automatic weight factor is displayed in Table II. We got improvised result after introducing property weight in our data integrator.

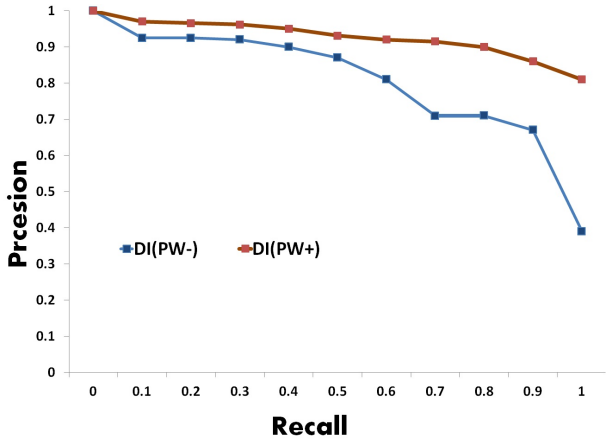


Fig. 2. Instance matching results against IIMB2009 where DI(PW+) and DI(PW-)are Data Integrator with Property Weight and without Property Weight (previous *aflood*) respectively.

Datasets	Trnasformation	Prec.	Rec.	F-Measure
001-010	Value transformations	1.00	1.00	1.000
011-019	Structural transformations	0.91	0.84	0.868
020-029	Logical transformations	1.00	1.00	1.000
030-037	Several combinations of the previous transformations	0.96	0.83	0.885

TABLE II. INSTANCE MATCHING RESULTS AGAINST IIMB BENCHMARKS WITH AUGMENTED WEIGHT FACTORS

B. Evaluation with IIMB2010 Data set

In OM-2010 instance matching track, the participants for IIMB2010 large dataset were Combinatorial Optimization for Data Integration (CODI) [12], Automated Semantic Mapping of Ontologies with Validation (ASMOV) [15] and RiMOM [11]. Fig. 3 shows the recall-precision graph of the participants[14] and the curves of our Data Integratort without property weight DI(PW-) and with property weigth DI(PW+). Our proposed method outperforms other methods in several cases although CODI shows better result in some cases.

VI. CONCLUSION AND FUTURE DIRECTION

In this paper, we present a novel and unique approach for weighting the property of data automatically. Our mathematical reasoning section discusses and logically satisfies the theoretical strength of the proposed method from different view points. Experiment and evaluation section exhibits how theoretically proven approach strongly contributes in achieving better outcome to integrate semantic data within same or among heterogeneous data sources. Our instance matcher with property weight provides better outcome than without property weight. Application of this method in other domain such as record linkage, entity resolution problem, identity recognition also help to evaluate its strength. Our future task covers to improve the time and space complexity of the proposed method. Finally, we would like to apply this generator in linked data project for integrating data semantically and efficiently.

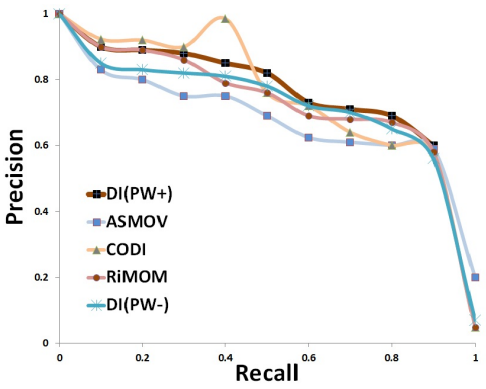


Fig. 3. Instance matching results against IIMB2010 where DI(PW+) and DI(PW-)are Data Integrator with Property Weight and without Property Weight respectively.

REFERENCES

[1] J. Aasman, "Allegro graph," Technical Report 1, Franz Incorporated, Tech. Rep., 2006.

[2] I. Fellegi and A. Sunter, "A Theory for Record Linkage," *Journal of the American Statistical Association*, vol. 64, no. 328, pp. 1183–1210, 1969.

[3] S. Castano, A. Ferrara, S. Montanelli, and D. Lorusso, "Instance matching for ontology population," in *SEBD*, 2008, pp. 121–132.

[4] K. Cortis, S. Scerri, I. Rivera, and S. Handschuh, "Discovering semantic equivalence of people behind online profiles."

[5] S. Araujo, J. Hidders, D. Schwabe, and A. de Vries, "Serimi-resource description similarity, rdf instance matching and interlinking," *Arxiv preprint arXiv:1107.1104*, 2011.

[6] W. Winkler, "The state of record linkage and current research problems," in *Statistical Research Division, US Census Bureau*. Citeseer, 1999.

[7] M. Seddiqui, S. Das, I. Ahmed, R. Nath, and M. Aono, "Augmentation of ontology instance matching by automatic weight generation," in *Information and Communication Technologies (WICT), 2011 World Congress on*. IEEE, 2011, pp. 1390–1395.

[8] R. Nath, H. Seddiqui, and M. Aono, "An efficient method for ontology instance matching."

[9] ——, "Resolving scalability issue to ontology instance matching in semantic web," in *Computer and Information Technology (ICCIT), 2012 15th International Conference on*. IEEE, 2012, pp. 396–404.

[10] H. Newcombe, *Handbook of record linkage: methods for health and statistical studies, administration, and business*. Oxford University Press, Inc., 1988.

[11] Z. Wang, X. Zhang, L. Hou, Y. Zhao, J. Li, Y. Qi, and J. Tang, "Rimom results for oaei 2010," *Ontology Matching*, vol. 195, 2010.

[12] J. Noessner and M. Niepert, "Codi: Combinatorial optimization for data integration–results for oaei 2010," *Ontology Matching*, p. 142, 2010.

[13] J. Euzenat, A. Ferrara, L. Hollink, A. Isaac, C. Joslyn, V. Malaisé, C. Meilicke, A. Nikolov, J. Pane, M. Sabou *et al.*, "Results of the ontology alignment evaluation initiative 2009," in *Proc. 4th ISWC workshop on ontology matching (OM)*, 2009, pp. 73–126.

[14] J. Euzenat, A. Ferrara, C. Meilicke, A. Nikolov, J. Pane, F. Scharffe, P. Shvaiko, H. Stuckenschmidt, O. Svb-Zamazal, V. Svtek, and C. Trojahn dos Santos, "Results of the ontology alignment evaluation initiative 2010," in *Proc. 5th ISWC workshop on ontology matching (OM), Shanghai (CN)*, P. Shvaiko, J. Euzenat, F. Giunchiglia, H. Stuckenschmidt, M. Mao, and I. Cruz, Eds., pp. 85–117.

[15] Y. R. Jean-Mary, E. P. Shironoshita, and M. R. Kabuka, "Asmov: Results for oaei 2010," *Ontology Matching*, vol. 126, 2010.

Evaluating Q-Learning Based Stateful Round Trip Time Estimation over High-Data-Rate Wireless Sensor Networks

A. B. M. Alim Al Islam[1,2] and Vijay Raghunathan[3]
[1,3]School of ECE, Purdue University, West Lafayette, IN 47907-2035, USA
[2]Department of CSE, Bangladesh University of Engineering and Technology, Dhaka-1000, Bangladesh
Email: [1]abmalima@purdue.edu, [2]alim_razi@cse.buet.ac.bd, [3]vr@purdue.edu

***Abstract*— Diversified applications of wireless sensor networks demand reliable data transmission, which generally experiences limited performance owing to the inherent resource constraints of sensor nodes. Estimated round trip time is one of the prominent aspects of reliable transmissions that possesses a potential to enhance the performance sustaining the resource constraints. Therefore, we have already proposed an efficient technique for estimating round trip time over wireless sensor networks. We have evaluated the proposed technique over low-data-rate networks, consisting low-bandwidth radios, considering classical deployments of sensor networks. However, recent deployments of wireless sensor networks have started to exploit high-bandwidth radios as they are now experiencing high-data-rate transmission. Therefore, in this paper, we evaluate the proposed estimation technique over high-data-rate wireless sensor networks consisting high-bandwidth radios. Our evaluation through both real testbed experiments and `ns-2` simulation reveals significant performance improvement using the proposed technique compared to that using other existing techniques.**

I. Introduction

Recent advanced applications of wireless sensor networks (WSNs) such as Internet connectivity [6] require reliable data transmission. These applications generally experience performance below expectation. One of the reasons for this happening is inherent resource constraints of sensor nodes. However, withstanding the notion of resource constraints, the paradigm of reliable data transmission itself appeals to be suitably adopted for WSNs. As TCP has already become the de facto standard for reliable data transmission, we attempt to suitably adopt its operational parameters for WSNs. TCP utilizes estimated Round Trip Time (RTT) as one of its operational parameters, which greatly influences network performance. Therefore, suitable estimation of RTT exhibits a great potential of enhancing the network performance. Consequently, we have proposed stateful Q-Learning based RTT estimation method [11], which we have termed as $QRTT$, for WSNs.

In the work presented in [11], we have mainly focused on low-data-rate (1 packet per several seconds or minutes) WSNs having low- bandwidth (250Kbps and 1Mbps) radios, as most of the classical applications of WSNs such as environmental monitoring and habitat monitoring experience low data transmission rate over low-bandwidth radios. However, as WSNs have matured over the past decade, more sophisticated networks have started to emerge. An example of such sophisticated networks is CeNSE [1] developed by Hewlett Packard, which aims building a "planet wide sensing network." One of its envisioned applications is to collect high-resolution data for seismic imaging to explore natural gas and oil reservoirs. Besides, structural heath monitoring [25], [17], [19], [18], acoustic sensing [9], and machine health monitoring [13] are a few other applications of the new generation of WSN applications. All these applications enable high-data-rate transmission over WSNs. Considering these applications, different approaches [22], [10] for developing WSN architectures start to consider deploying high-bandwidth radios in WSNs. Therefore, it becomes of the utmost importance to analyze whether we could exploit the notion of $QRTT$ in high-data-rate (one or tens of packets per second) WSNs having high-bandwidth (such as 54Mbps) radios.

Consequently, we experimentally analyze the impact of utilizing $QRTT$ in high-data-rate WSNs in this paper. Here, we make the following contributions: 1) We evaluate the performance of $QRTT$ for high-data-rate transmission over a real testbed consisting high-bandwidth radios. In the evaluation, we compare the performance of $QRTT$ against that of two other RTT estimation techniques in literature: Jacobson's RTT [12] and Fallon's RTT [8]. Here, we find up to 37% increased network throughput and 31% decreased average energy per bit using $QRTT$ compared to that using other two techniques. 2) Besides, we perform rigorous `ns-2` simulation for the same comparison over a variety of different network settings. Here, we find up to 245% increased network throughput and 60% decreased average energy per bit using $QRTT$ compared to that using other techniques.

II. Overview of Round Trip Time Estimation Mechanism in QRTT

$QRTT$ separately considers successful transmission and packet drops over wireless medium by adopting two states for data transmission: $success$ and $failure$. Receiving an acknowledgment from a receiver indicates a successful transmission, which moves a sender to $success$ (S) state. Besides, expiring retransmission timer indicates a packet drop, which moves a sender to $failure$ (F) state (see Fig. 1, where P_{xy} denotes the probability of transition from state x to state y). For estimating RTT in each state, $QRTT$ exploits a modified form of original equation [14] of Q-Learning as [11]

$$Q(s_t) \leftarrow Q(s_t) + \alpha_{s_t} \times \{d_{t+1} + \gamma_{s_{t+1}} \times Q(s_{t+1}) - Q(s_t)\} \quad (1)$$

Here, t and $t+1$ indicate previous and current time indices. Besides, s_t, α_{s_t}, and d_t are state, learning rate, and measured delay at time t respectively. In addition, γ denotes a discount factor which varies over two different states. Using these parameters, $QRTT$ estimates RTTs in the two different states using a one-dimensional matrix Q. For further elaboration

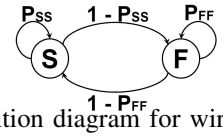


Fig. 1: State transition diagram for wireless transmissions

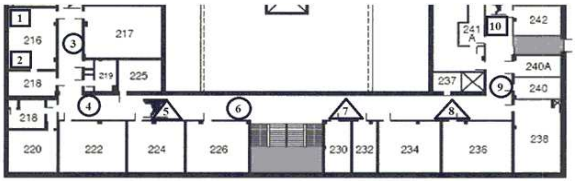


Fig. 2: Node placements in testbed (rectangles denote source nodes, triangles denote destination nodes, and circles denote intermediate nodes)

of the estimation technique and for choices of its parameter values, please refer to [11].

III. Testbed Evaluation of QRTT

We evaluate the performance of $QRTT$ over a real testbed. In our evaluation, we compare the performance of $QRTT$ against that of Jacobson's RTT [12] and Fallon's RTT [8]. Here, Jacobson's RTT represents a classical estimation approach whereas Fallon's RTT represents a specialized estimation approach solely intended for wireless networks. In our evaluation, we choose TCP Retransmission Timeout (RTO) adjustment as the underlying application of RTT estimation, as RTO adjustment is the most prominent one among all applications of RTT estimation. Before presenting outcomes of such evaluation, we briefly discuss our testbed settings in the next subsection.

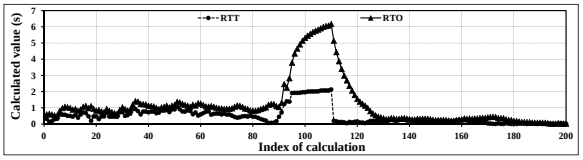
A. Testbed Settings

In our testbed, we deploy ten wireless nodes in the 2nd floor of EE building at Purdue University (Fig. 2). Each node in our testbed is a Beagleboard [4] having one 54Mbps Trendnet 802.11 dongle (TEW-648UB) [16] connected to it. In these nodes, we utilize Ubuntu as the operating system. We enable three multi-hop flows ($node1 \rightarrow node7$, $node2 \rightarrow node8$, and $node10 \rightarrow node5$) over the nodes. We enable TCP Westwood [3] (a variant intended for wireless networks) as the transport layer protocol and Olsrd [24] as the routing daemon to determine routes for the multi-hop flows. Here, we assign different data rates over the flows. Over the first two flows, i.e., at source node1 and at source node2, we enable a data transmission rate of 50 packets per second (pps). Over the last flow, i.e., at source node10, we enable a data transmission rate of 100 pps. Here, each packet consists an application payload of 1KB for all flows. We take measurements after enabling all flows over six iterations each operating for 25 seconds with 10 seconds intermittent delay between two successive iterations.

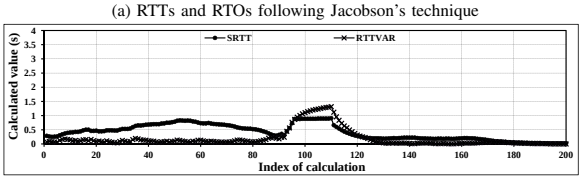
B. Analysis over Estimated RTTs

Now, we present our testbed evaluation results. First, we analyze the variations over RTTs, and thus in turn over RTOs, following different estimation techniques. Fig. 3 and Fig. 4 show variations in RTTs, RTOs, and components of RTOs for the three RTT estimation techniques over 200 successive values captured from a node in our testbed.

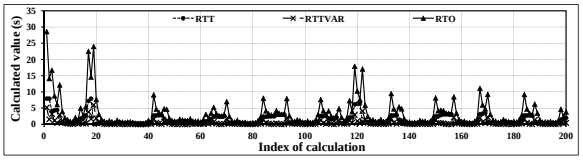
In the case of Jacobson's technique, there might be severe increases in RTOs in response to moderate changes in RTTs. In



(a) RTTs and RTOs following Jacobson's technique

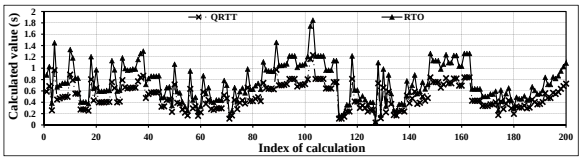


(b) Components of RTOs (SRTTs [12] and RTTVARs [12]) following Jacobson's technique

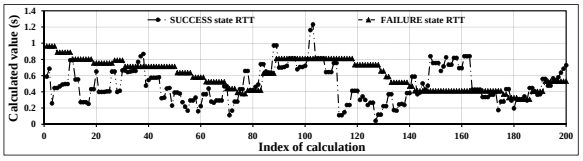


(c) RTTs, RTTVARs [8], and RTOs following Fallon's technique

Fig. 3: RTOs and their components for Jacobson's RTT and Fallon's RTT



(a) $QRTT$s and RTOs following stateful Q-Learning

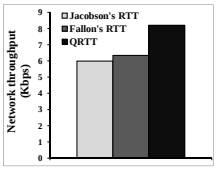


(b) Components of $QRTT$s (RTTs in two different states)

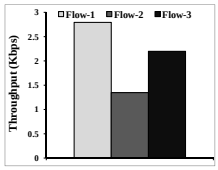
Fig. 4: RTOs and their components for $QRTT$

Fig. 3a, the RTOs corresponding the indices $90 \sim 130$ present an example of such severe increases. We can investigate the reason behind such changes by analyzing the changes in the components of RTOs in Jacobson's technique, i.e., SRTT (smoothed RTT that takes a weighted average of previously experienced RTTs) and RTTVAR (variance over the experienced RTTs). Fig. 3b shows that SRTT follows the changes in RTT during the period of the severe increases, whereas RTTVAR attains significant values during the period. The significant values of RTTVAR mainly contribute to severe increases in RTOs. Such sever increases result in over-adjustments of RTOs during a significant amount of time. These over-adjustments result in unnecessary waiting periods in case of failed transmission attempts. Such unnecessary waiting periods, in turn, reduce the number of successfully transmitted bits in the network, i.e., the network throughput.

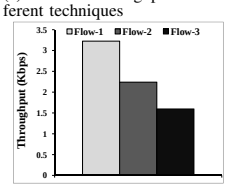
Next, in the case of Fallon's technique, similar over-adjustments arise in a different manner. Here, the over-adjustments occur in a sporadic manner in contrast to the scenario over a long duration in the case of Jacobson's technique. Fig. 3c shows such sporadic over-adjustments that exhibit significantly high values for RTOs ($\geq$ 5s), and thus lower network throughput in a similar manner described earlier



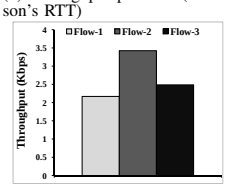
(a) Network throughput for different techniques



(b) Throughput per flow (Jacobson's RTT)



(c) Throughput per flow (Fallon's RTT)



(d) Throughput per flow ($QRTT$)

Fig. 5: Network performance of different RTT estimation techniques in terms of throughput

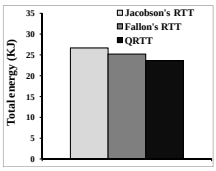
in the case of Jacobson's technique.

In the case of our proposed $QRTT$, RTOs almost always closely follow RTTs eliminating the over-adjustments that we have found with other two techniques. Fig. 4a shows the closely following nature of RTOs. Here, the closely following nature results from two different estimations of RTTs in $success$ and $failure$ states, which we present in Fig. 4b. These close adjustments of RTOs result in an improved network performance that we elaborate next.

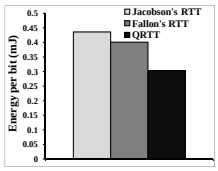
C. Network Performance Analysis

We start our analysis on network performance through focusing on throughput. Fig. 5 shows the performance with different RTT estimation techniques. Here, Fig. 5a shows that we can significantly boost up network throughput using $QRTT$ in comparison to both using Jacobson's technique (by 37%) and using Fallon's technique (by 29%). Further, we also show per flow throughput for the three techniques in Fig. 5b, Fig. 5c, and Fig. 5d respectively. Here, the standard deviation over the per flow throughput using $QRTT$ is less than that using Jacobson's technique and Fallon's technique by 10% and 20%. The decreased value in the standard deviation implies higher extent of fairness using $QRTT$ compared to the other alternatives.

The analysis on throughput reveals that using $QRTT$ significantly increases the number of successfully transmitted bits in the network. In another way, using $QRTT$ requires less amount of time for transmitting same volume of data, which should in turn require less amount of energy. To prove this rationale, we perform an analysis over energy consumption over the network. Here, we consider only the energy consumption by radios as the radios consume most of the energy in a wireless network [7]. We start the analysis by focusing on total energy consumption in the network using different RTT estimation techniques, which we show in Fig. 6a. This figure reveals that using $QRTT$ consumes less total energy compared to that consumed using both Jacobson's technique and Fallon's technique. The lower energy consumption in accordance with higher network throughput, which we have already presented, results in a significant decrease in average energy consumption per transmitted bit in the case of using $QRTT$ as presented in Fig. 6b. Here, using $QRTT$ achieves 31% and 25% decreased

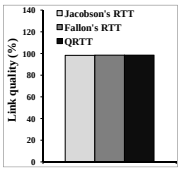


(a) Total energy

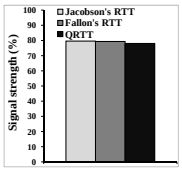


(b) Energy per bit

Fig. 6: Energy efficiencies of three estimation techniques

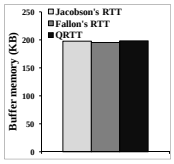


(a) Link qualities

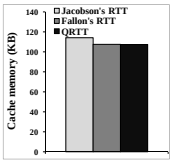


(b) Signal strengths

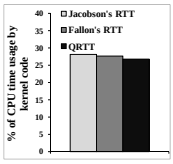
Fig. 7: Network conditions during testbed experiments



(a) Buffer usages



(b) Cache usages



(c) CPU usages

Fig. 8: Memory and CPU usages by different techniques

average energy per bit compared to using Jacobson's technique and using Fallon's technique respectively.

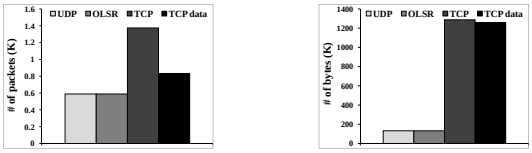
D. Network Conditions and Operational Overheads Analysis

Now, we will analyze network conditions and operational overhead observed during our experiment. Here, we analyze network conditions to confirm that all experiments over our testbed are conducted in similar environments. We utilize two metrics for analyzing the network conditions: link quality and signal strength. We capture their values by taking readings at node6, which is placed almost at the middle of our testbed. We present average values of captured link qualities and average signal strengths in Fig. 7. Here, Fig. 7a and Fig. 7b demonstrate that average link quality and average signal strength exhibit almost similar values during our experiments with all RTT estimation techniques.

Next, we analyze system overheads incurred by different RTT estimation techniques (Fig. 8). We perform the analysis in term of memory and CPU usages (captured from node1) by the techniques. The goal of this analysis is to investigate system requirements for $QRTT$ compared to that required by other estimation techniques.

In the case of analyzing memory usages, we focus on both buffer (used while data is being *read from* the disk) and cache (used while data is being *written to* the disk) memories. $QRTT$ incurs similar buffer memory usage compared to that incurred by other two techniques (Fig. 8a), and similar cache memory usage compared to Fallon's technique requiring 6% less memory compared to Jacobson's technique (Fig. 8b). Further, we also analyze CPU usages by all the techniques (Fig. 8c). Here, due to the simple calculation in $QRTT$, it incurs less CPU usage (by 5% and 3% respectively) compared to other two techniques.

Finally, we analyze overhead associated with the routing algorithm utilized in our testbed experiment. We perform this



(a) Number of packets (b) Number of bytes

Fig. 9: Network usages by different types of data flows

analysis to reveal whether the Olsr daemon, which we use for multi-hop routing over our testbed, results in any imposing overhead in the experiments. For such analysis, we utilize Wireshark [5] to capture network traffic in the testbed. We present summary of such capturing in Fig. 9. Here, the routing module, i.e., Olsrd, exhibits 30% network overhead in terms of the number of packets transmitted over the testbed (Fig. 9a). However, it incurs only 9% network overhead in terms of the number of bytes (Fig. 9b). Therefore, we can state that the routing module does not incur an imposing overhead over the testbed.

Up to this point, we evaluate the performance of $QRTT$ using real testbed experiments. Here, the number of nodes is limited to a small value of 10. However, in practice, the number of nodes may become significantly larger compared to the number adopted in our testbed. Therefore, we should also evaluate the performance of $QRTT$ over networks having large numbers of nodes. To perform such evaluation, we simulate $QRTT$ in the following section.

IV. Evaluation Using Simulation

We evaluate the efficacy of $QRTT$ by simulating various network topologies in this section. Here, we choose `ns-2` as our simulation platform considering its wide acceptability in wireless research. We have already presented simulation results for low-bandwidth radio and low-data-rate applications in [11]. In this section, we present simulation results pertinent for high-bandwidth radio and high-date-rate applications. Before presenting the simulation results, we illustrate common network settings.

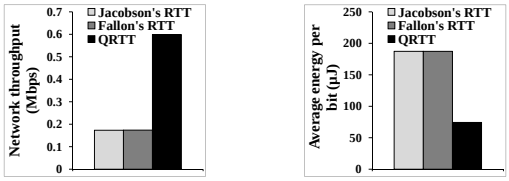
A. Network Settings

We choose 54Mbps 802.11g as the high-bandwidth radio in our simulation. We utilize a transmission range of 140m and a carrier sensing range of 280m for the radio. Besides, we utilize the measurements provided in [23] to calculate radio power consumption. Here, we compute power consumptions at different radio states, i.e., transmission, reception, idle, and sleep states, as we consider duty-cycling [15] in the simulation.

Besides, we use a droptail priority queue having a capacity of 50 packets at each node for storing packets on transit. In addition, we adopt two-ray ground reflection model [2] as the radio propagation model, DSDV [20] as the routing layer protocol, and TCP Tahoe [21] as the transport layer protocol. With these settings, we conduct simulation for a duration of 75 seconds with an initial period of 50 seconds for routing path establishment.

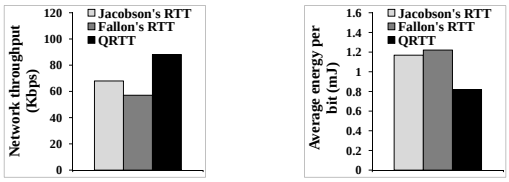
B. Simulation Results

We start our evaluation of $QRTT$ over a grid topology having 10×10 nodes. The nodes are uniformly placed at



(a) Network throughput (b) Energy per bit

Fig. 10: Performance over a grid topology



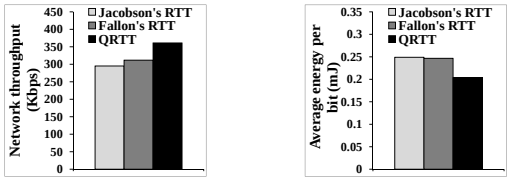
(a) Network throughput (b) Energy per bit

Fig. 11: Performance over a random topology for 1pps data rate and 1KB application data

100 meters away from adjacent nodes in the same row and in the same column. We conduct simulation with 20 end-to-end cross flows, where each flow exhibits a data rate of 1 pps with 1KB application data. Fig. 10 presents the performances of three RTT estimation techniques over the grid topology. $QRTT$ achieves significant improvement in both average network throughput and average energy per bit in comparison to Jacobson's RTT and Fallon's RTT in this topology. The improvement in network throughput are 245% and 244% accordingly (Fig. 10a), and the improvement in average energy per bit is 60% (Fig. 10b) over other two techniques.

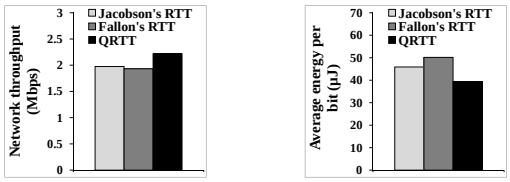
In addition to the grid topology, we also evaluate $QRTT$ over random topologies. For all the random topologies, we present average of 30 simulation runs. We start the evaluation over random topologies with 100 nodes within an area of 1km^2. First, we evaluate the random topology with 20 random flows, where each flow exhibits a data rate of 1pps with 1KB application data. Fig. 11 shows the performances of three estimation mechanisms for this setting. Fig. 11a reveals that $QRTT$ achieves 30% and 54% increased network throughput in comparison to Jacobson's RTT and Fallon's RTT respectively. Besides, Fig. 11b shows that $QRTT$ decreases average energy per bit by 30% and 33% in comparison to other two estimation techniques accordingly.

In our evaluation of $QRTT$ over high-data-rate networks, we also focus on the impacts of variations in application data size and data rate over the performance of $QRTT$. Therefore, we perform additional two simulations: one with increased application data size and another with increased data rate. Fig. 12 depicts the performances of three estimation techniques with an increased application data size of 5KB. Here, $QRTT$ achieves 22% and 16% increased network throughput in comparison to Jacobson's and Fallon's estimation techniques accordingly (Fig. 12a). Corresponding improvements in average energy per bit are 18% and 17% respectively (Fig. 12b). Besides, we also simulate the random topology with an increased data rate of 50 pps (Fig. 13). Here, $QRTT$ achieves 12% and 15% increased network throughput in comparison



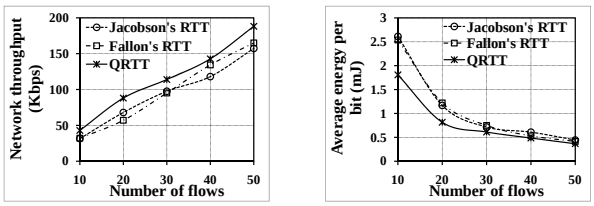
(a) Network throughput (b) Energy per bit

Fig. 12: Performance over a random topology for 1pps data rate and 5KB application data



(a) Network throughput (b) Energy per bit

Fig. 13: Performance over a random topology for 50pps data rate and 1KB application data



(a) Network throughput (b) Energy per bit

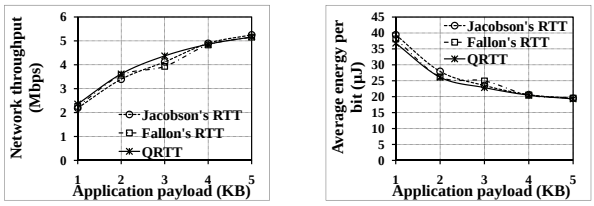
Fig. 14: Performance over a random topology for varying number of random flows over the network

to Jacobson's and Fallon's estimation techniques accordingly (Fig. 13a). Corresponding improvements in average energy per bit are 14% and 22% respectively (Fig. 13b).

In addition, we also attempt to evaluate the change in performance of $QRTT$ with a variation in the number of random flows. Therefore, we vary the number of random flows from 10 to 50 where each flow exhibits a data rate of 1pps with 1KB application data (Fig. 14). Here, $QRTT$ achieves an average of 22% and 19% improvement in network throughput along with 26% and 25% improvement in average energy per bit compared to other two techniques accordingly.

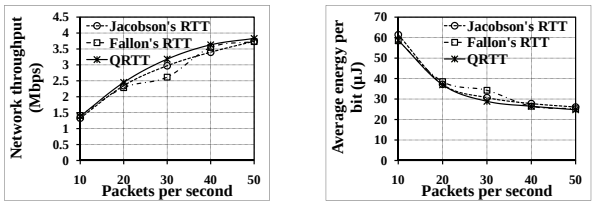
Comparing the performance improvement in the first random setting against the performance improvement in subsequent random settings with increased data size, data rate, and number of flows, we find that the extent of improvement decreases in the subsequent cases. Here, the subsequent cases experience presence of a large number of simultaneous transmissions over the network. Therefore, to identify the trend in performance of $QRTT$ in presence of large number of simultaneous transmissions, we further simulate two different scenarios: with varying application data size and with varying data rates over a large number of flows.

In the first scenario, we vary application data size from 1KB to 5KB with 50 random flows each having a data rate of 20pps (Fig. 15). Here, the improvement in network throughput achieved by $QRTT$ gracefully decreases with an increase in application data size (Fig. 15a). $QRTT$ retains a comparable network throughput even with a large application data size such as 5KB. Besides, we find a similar outcome for average



(a) Network throughput (b) Energy per bit

Fig. 15: Performance over a random topology for varying size of application data



(a) Network throughput (b) Energy per bit

Fig. 16: Performance over a random topology for varying data rate

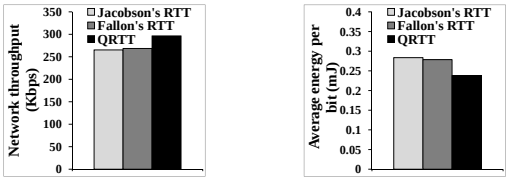
energy per bit with the increase in application data size (Fig. 15b). In both the cases, we find small improvement using $QRTT$. Next, in case of the second scenario, we vary data rate from 10pps to 50pps with 50 random flows with an application data size of 1KB (Fig. 16). Similar to the previous case, both network throughput (Fig. 16a) and average energy per bit (Fig. 16b) of $QRTT$ remains comparable, exhibiting small improvements, compared to that of other mechanisms irrespective of the increase in data rate.

It is worth mentioning that the performance of $QRTT$ exhibits smooth changes with the variations in application data size and data rate in the previous two cases even though the performances of other two estimation mechanisms exhibit slight fluctuations. The reason behind the smooth changes is the highly stable RTT estimations in two different states in case of $QRTT$.

Up to this point, we evaluate the performance of $QRTT$ with either fixed cross flows or with random flows. These types of flows adequately characterize the performance of $QRTT$ over ad-hoc connectivity. Next, we evaluate the performance of $QRTT$ over gateway-oriented flows to characterize the performance over infrastructure connectivity.

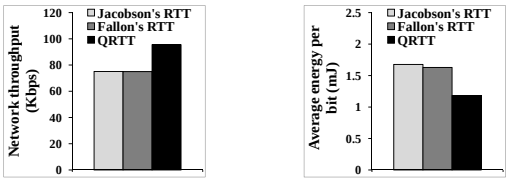
We start with an infrastructure network topology having 100 nodes randomly placed within 1km^2 coverage area along with a gateway placed at the center of the coverage area (at (500m, 500m), to be more specific). We enable 25 flows from random source nodes to the gateway and 25 flows from the gateway to random destination nodes. Each flow transmits packets at a data rate of 1pps having 1KB application data. Fig. 17 shows the performances of three estimation mechanisms in this setting. Here, $QRTT$ achieves 12% and 10% improvement in network throughput in comparison to Jacobson's and Fallon's estimation techniques respectively (Fig. 17a). Besides, the improvement in average energy per bit are 16% and 15% respectively (Fig. 17b).

In addition, we also simulate a similar infrastructure topology except we place the gateway at a corner of the coverage



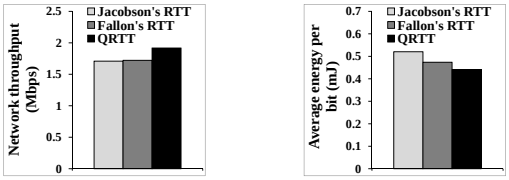
(a) Network throughput (b) Energy per bit

Fig. 17: Performance over a random topology having a sink at the center of coverage area



(a) Network throughput (b) Energy per bit

Fig. 18: Performance over a random topology having a sink at a corner of coverage area



(a) Network throughput (b) Energy per bit

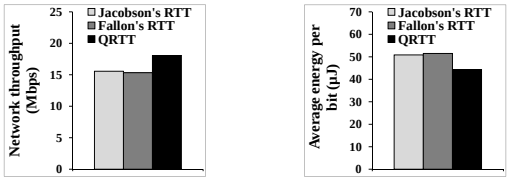
Fig. 19: Performance over a large-scale ad-hoc network

area (at (1km, 1km), to be more specific). Fig. 18 depicts the performances of three estimation mechanisms in this topology. Here, $QRTT$ achieves 27% improved network throughput compared to both of other two techniques (Fig. 18a). In addition, the improvement by $QRTT$ in average energy per bit are 29% and 27% compared to Jacobson's and Fallon's techniques respectively (Fig. 18b).

Investigation up to this point reveals different aspects of $QRTT$ performance in small-scale deployment. However, recent emergence of large-scale deployments of WSNs such as CeNSE [1] necessitates evaluating the performance in large-scale network topologies. Therefore, we also focus on large-scale evaluation of $QRTT$.

We evaluate the performance of $QRTT$ in two types of large-scale deployment: ad-hoc and infrastructure. In both cases, we consider 1000 randomly deployed sensor nodes within a coverage area of 10km^2. In the ad-hoc deployment, we enable 100 random flows having 50pps data rate with 1KB application data. Fig. 19 shows the performances of three estimation techniques in this case. Here, $QRTT$ achieves 12% and 11% improvement in network throughput (Fig. 19a), and 15% and 7% improvement in average energy per bit (Fig. 19b) compared to Jacobson's and Fallon's estimation respectively.

In the infrastructure deployment, we consider random placement of 20 gateways over the network. Here, we enable 50 flows from random source nodes to their nearest gateways and 50 flows to random destination nodes from their nearest gateways. Each flow possesses 50pps data rate with 1KB application data. Fig. 20 shows the performances of three estimation techniques in the infrastructure deployment. Here,



(a) Network throughput (b) Energy per bit

Fig. 20: Performance over a large-scale infrastructure network having gateways

$QRTT$ achieves 16% and 18% improvement in network throughput (Fig. 20a), and 13% and 14% improvement in average energy per bit (Fig. 20b) compared to Jacobson's and Fallon's estimation mechanisms respectively.

V. Conclusion

WSNs have started experiencing utilization of high-bandwidth radios to support high-data-rate transmission. The performance of $QRTT$, a Q-Learning based stateful RTT estimation technique specifically devised for WSNs, is yet to be evaluated over such networks. Therefore, in this paper, we have evaluated the the performance of $QRTT$ over high-data-rate WSNs. Our evaluation covers both testbed experiment and `ns-2` simulation over a variety of network settings. Our evaluation results demonstrate that we can achieve significant performance improvement using $QRTT$ compared to that using other state-of-the-art techniques.

References

[1] CeNSE, http://www.hpl.hp.com/news/2009/oct-dec/cense .html.
[2] Two-ray ground reflection model. http://www.isi.edu/nsnam /ns/doc/node-219.html.
[3] C. Casetti, M. Gerla, S. Mascolo, M. Y. Sanadidi, and R. Wang. Tcp westwood: end-to-end congestion control for wired/wireless networks. *Wireless Networks*, 8(5):467–479, September 2002.
[4] G. Coley. Beagleboard system reference manual, 2009.
[5] G. Combs et al. Wireshark-network protocol analyzer. 2008.
[6] P.A.C. da Silva Neves and J.J.P.C. Rodrigues. Internet Protocol over Wireless Sensor Networks, from Myth to Reality. *Journal of Communications*, 5.
[7] P. Dutta, D. Culler, and S. Shenker. Procrastination might lead to a longer and more useful life. *Proceedings of HotNets-VI*, 2007.
[8] S. Fallon, P. Jacob, Y. Qiao, L. Murphy, E. Fallon, and A. Hanley. An Analysis of Alterations to the SCTP RTO Calculation Mechanism for WLAN Environments. *Wireless and Mobile Networking*, pages 95–108, 2008.
[9] L. Girod, M. Lukac, V. Trifa, and D. Estrin. The design and implementation of a self-calibrating distributed acoustic sensing platform. In *Proceedings of Sensys*, pages 71–84. ACM, 2006.
[10] A.B.M.A.A. Islam, M.S. Hossain, V. Raghunathan, and Y.C. Hu. Backpacking: Deployment of heterogeneous radios in high data rate sensor networks. In *Proceedings of ICCCN*, pages 1–8. IEEE, 2011.
[11] A.B.M.A.A. Islam and V. Raghunathan. QRTT: Stateful Round Trip Time Estimation for Wireless Embedded Systems Using Q-Learning. *Embedded Systems Letters (ESL)*, 4(4):102–105, 2012.
[12] V. Jacobson. Congestion avoidance and control. *ACM SIGCOMM Computer Communication Review*, 18(4):314–329, August 1988.
[13] D. Jarrell, D. Sisk, and L. Bond. Prognostics and condition based maintenance (cbm) - a scientific crystal ball. In *Proceedings of ICAPP*, June 2002.
[14] L.P. Kaelbling, M.L. Littman, and A.W. Moore. Reinforcement learning: A survey. *Arxiv preprint cs/9605103*, 1996.
[15] P. Levis, S. Madden, J. Polastre, R. Szewczyk, K. Whitehouse, A. Woo, D. Gay, J. Hill, M. Welsh, E. Brewer, et al. Tinyos: An operating system for sensor networks. *Ambient Intelligence*, pages 115–148, 2005.
[16] Mini Wireless N USB Adapter. *http://www.trendnet.com/ products/proddetail.asp?prod=190_TEW-648UB&cat=175*.
[17] NetSHM: Project Overview, http://enl.usc.edu/projects /netshm/.
[18] J. Paek, K. Chintalapudi, R. Govindan, J. Caffrey, and S. Masri. A wireless sensor network for structural health monitoring: Performance and experience. In *Proceedings of EmNetS-II*, pages 30–31. Citeseer, 2005.
[19] J. Paek, N. Kothari, K. Chintalapudi, S. Rangwala, and R. Govindan. The performance of a wireless sensor network for structural health monitoring. 2004.
[20] C.E. Perkins and P. Bhagwat. Highly dynamic Destination-Sequenced Distance-Vector routing (DSDV) for mobile computers. In *Proceedings of SIGCOMM*, pages 234–244, August 1994.
[21] L.L. Peterson and B.S. Davie. *Computer networks: a systems approach*. Morgan Kaufmann Publishers Inc., San Francisco, CA, 2007.
[22] C. Sengul, M. Bakht, A. F. Harris, T. Abdelzaher, and R. Kravets. Improving energy conservation using bulk transmission over high-power radios in sensor networks. In *ICDCS*, pages 801–808. IEEE, 2008.
[23] T. Stathopoulos, M. Lukac, D. Mclntire, J. Heidemann, D. Estrin, and W. J. Kaiser. End-to-end routing for dual-radio sensor networks. In *Proceedings of INFOCOM*, pages 2252–2260. IEEE, 2007.
[24] A. Tonnesen, T. Lopatic, H. Gredler, B. Petrovitsch, A. Kaplan, and SO Tcke. Olsrd: an ad hoc wireless mesh routing daemon, 2008.
[25] N. Xu, S. Rangwala, K.K. Chintalapudi, D. Ganesan, A. Broad, R. Govindan, and D. Estrin. A wireless sensor network for structural monitoring. In *Proceedings of Sensys*, pages 13–24. ACM, 2004.

Design and Implementation of a RF Controlled Robotic Environmental Survey Assistant System

[1]Md. Shamsul Alam, [2]Insan Arafat Jamil, [3]Khizir Mahmud and [4]Najmul Islam

[1, 2, 3] Department of Electrical & Electronic Engineering, Chittagong University of Engineering & Technology, Bangladesh.
[4]Department of Electrical Engineering, Northwestern Polytechnical University, P. R. China
[1]samrat.ieee@yahoo.com, [2]iaj.ankon@gmail.com, [3]khizirbd@gmail.com and [4]najmulislam@ovi.com

Abstract— **Survey operation to collect information about environmental parameters is an inevitable part in various cases. But manual data logging by human is dangerous in perilous places. This paper deals with design and implementation of a RF Controlled Robotic Environmental Survey Assistant System for remote survey operations, which can assist in remote data acquisition of environmental parameters like temperature, humidity and the presence and level of LPG gas in the air. Technically, two individual units make up the complete survey assistant system. First one is the wireless control unit for controlling and monitoring the survey process, whereas the second one is a remote controlled data collector robot unit which executes the survey process. The survey robot is equipped with sensor units required for measurement of temperature, humidity and combustible gas level as environmental parameters of the survey zone, ultrasonic sensor for obstacle detection in the path of the robot and GPS receiver for collecting information about position of the robot. A Radio Frequency (RF) Amplitude Shift Keying (ASK) transceiver channel enables the communication between the robot and control panel. The robot can be operated in two different modes. In autonomous mode, the robot travels a random path depending on obstacle avoidance technique and stores or collects and broadcasts live data stream to the control panel. Each time the robot is set to autonomous mode, after completion of a limited time survey operation, the robot comes back to it's starting position where autonomous mode was started. Obstacle sensor units installed in the robot helps it to detect and avoid obstacles in its path. The real-time data visualization is achieved on a Liquid Crystal Display (LCD) in the control panel and the mode of operation can be select from the control panel by operator.**

Index Terms— **Control unit, Environmental parameters, Manual and autonomous mode, RF ASK transceiver, Survey unit**

I. INTRODUCTION

Survey operations often become necessary to collect information about environmental parameters for various purposes. Most of the cases, the survey operations involve manual data logging by direct human effort to achieve greater accuracy and reliability. But sometimes it becomes dangerous for mankind to be present physically in the area of survey operation. Survey operation in risky places like old rejected mine or in radio-active zones can be dangerous for physical presence of human being. So nowadays human are being replaced by robotic system in challenging works and risky environment. This paper depicts a model of survey operation system with the help of a remote controlled robot which can perform a set of survey in a remote place and broadcasts the information to its control unit. The environmental survey assistant systems can be used inside collapsed buildings, radioactive zones, mines etc. In this system, humidity, temperature [2] and combustible gas level are selected as survey parameters. There are several existing systems which are capable of performing such survey activities with relatively complex and costly equipment. Though having capability of a long range communication, these systems often become expensive for light survey operations. This project work was intended to derive a cost effective remote survey operation system with a moderate communication range (100 meters-from datasheet) using locally available components. The system comprises of two different units- the wireless control unit and the robot unit equipped with a set of sensors and wireless digital communication module. Capability of communicating with multiple devices provides facility of controlling multiple robots simultaneously through a single control unit. Besides, the proposed system offers two different modes of operation- autonomous and manual, which inturns makes the system more useful and dynamic when subjected to lengthy operation time, or in the radio dropout zones.

II. METHODOLOGY

The system mainly consists of a wireless control unit and one (or more) robot unit for survey operation in remote areas. The wireless control unit acts as the eye of the operator while operating the survey mission. A live wireless connection between the control unit and the robot unit/units can be established using modern low power digital RF transceiver device NRF24L01P, integrated in both the units of the system. This module offers smooth wireless data access up to 100 meters of range, with 2.4 GHz carrier frequency and a maximum of 2Mbps data rate. A LCD display serves as the visual interface for the operator, in which the command summary and response of the remote unit are displayed. Operator gives the appropriate command to the robot

using a keypad interfaced with the control unit. Keypad in the control units provides individual buttons for individual movements of the robot in addition with additional control buttons for changing of modes, acquiring data etc. Microchip's midrange PIC microcontroller serves as the processor of the control unit, which runs the installed firmware at 20Hz clock rate. LCD and keypad in the unit are accessed through built in digital input output module of the microcontroller. The RF transceiver NRF24L01P is accessed by built in MSSP module of the microcontroller using SPI (Serial Peripheral Interface) communication protocol. The microcontroller accesses individual control and data registers and read from / write to them according to implemented sequence for utilization of the module.

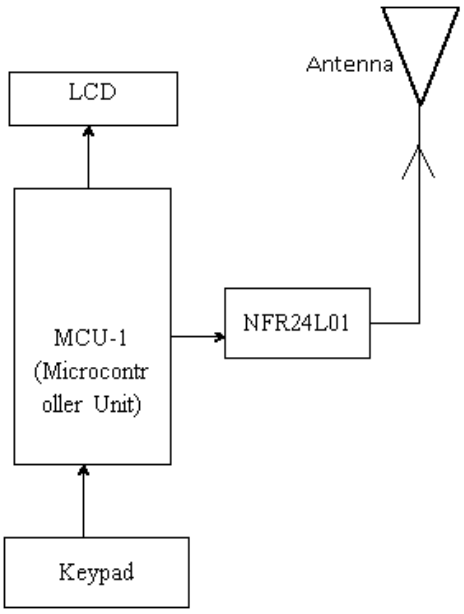


Fig. 1: Block diagram of Wireless Control Unit

The robot unit / remote survey unit acts as the core of system, responsible for appropriate execution of the survey operation. This unit is equipped with a simple land rover for exploring the survey area, a wireless transceiver module NRF24L01P for communicating with the remote control unit, a GPS device for global positioning, humidity, temperature and LPG gas sensors and related instrumentation circuits for measuring environmental parameters, an ultrasonic sensor for detection of obstacles in it's path and a central processing unit based on midrange PIC microcontroller. There are two different operation modes for this unit- the autonomous mode and the manual mode. The robot switches between these modes by the commands of the operator. In case of long time communication breakdown, the robot automatically switches to autonomous mode, performs a random survey operation throughout the region, and stores them into memory and returns back to its primary position using GPS. In the manual mode, the robot operates according to the commands received from the control unit.

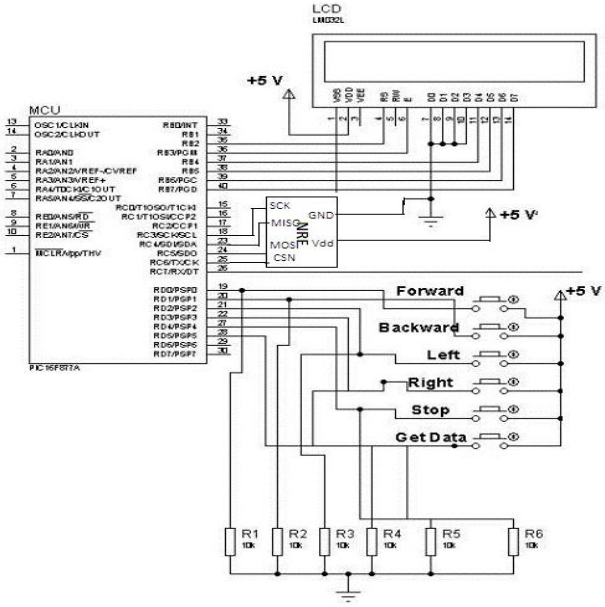


Fig. 2: Circuit diagram for wireless control unit

The processor of the robot unit continuously monitors the data packets received from the control unit, classifies them and executes required operation for every meaningful word. Words are mainly classified as control words (for controlling the movement of the robot), acquisition word (for starting remote data acquisition process) and mode change word (for changing the mode between manual and autonomous). While in manual mode, the robot continuously streams it's position to the control unit which it obtains from the GPS receiver, thus assists the operator of the robot deciding the next step of operation. A 12V battery pack inside the robot structure provides the electrical energy required for a complete operation.

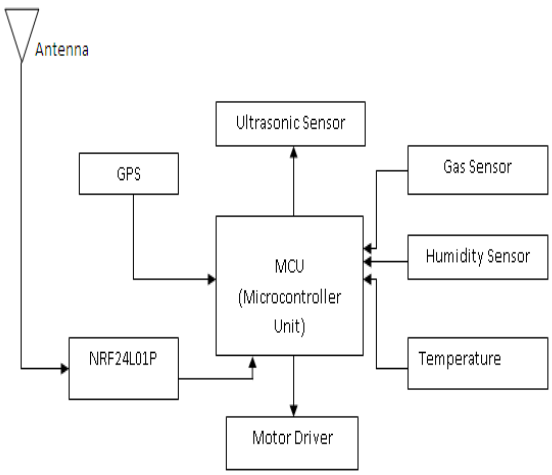


Fig. 3: Block Diagram of Remote Survey Unit

III. SENSING OF PARAMETERS FOR SURVEY OPERATION

Sensors play a cardinal part of this project implementation. To acquire information about environment [6] parameters, humidity sensor, temperature sensor and LPG gas sensors have been used. Responses of the sensors are conditioned properly for digital reading using appropriate signal conditioning circuits and fed to the internal A/D converter module of the microcontroller for reading in a programmed manner.

A. Sensing of temperature

For the temperature sensing LM35 temperature sensor has been used in temperature measurement circuitry. This kind of temperature sensor has an advantage for its wide span of temperature sensing both in positive and negative range. It can sense from -55 °C to the 150 °C [3][7] and hence a smart choice for its long range of temperature sensing and accurate response as well. Output voltage of this sensor is proportional to the temperature of the sensor surface with scaling factor of 10 mV/ °C [4].This circuit setup will measure the positive temperature from 0°C to 155°C. An internal analogue to digital converter (ADC) of the PIC16F77A microcontroller directly converts the sensor output voltage to 10 bit digital representation [1] [3]. ADC voltage measurement range is from 0 to 5 V [4][8].

As the reference voltage, V_{ref} is 5V and ADC is 10 bit, the output voltage is mapped between 0 to 1023 for 0V and 5V respectively.

The resolution of A/D conversion is as following Resolution = 5/1024 = 0.004883 V/Count.

So, corresponding to any input voltage digital output is $V_{out} = V_{in}/0.001168$. Finally, doing all the conversion and process the temperature can be found back from the 10-bit digital number. So need to divide V_{out} by 10mV. Therefore, temperature = $V_{out}/10m = 100V_{out}$. The sensor V_{out} pin is connected to AN0 pin of PIC 16F877A [9].

B. Sensing of humidity

Humidity Sensor Module (HSM) 20G is used in this system for humidity measurement. The sensor has two output pin of which one for humidity and another for temperature. The output signal comes in the form of analogue signal which is read by built in A/D converter module of PIC microcontroller and processed this value for further calculation. Calculation of humidity value follows the equation-

Relative humidity,

$$h = (3.710\times V_h^3) - (20.65\times V_h^2) - (64.8\times V_h) - 27.43$$

Where, h= humidity of the environment, V_h = output voltage of HSM-20G sensor.

The humidity equation has been derived from its characteristic curve [5].

C. Sensing of LPG gas level

To know the information about combustible gas existence in the survey area, MQ-5 gas sensor is used in this system considering its fast response, stability and life time. It shows high sensitivity to light petroleum gas, natural gas, town gas whereas shows less sensitivity to alcohol and smoke. The enveloped MQ-5 have 6 pins, 4 of them used to fetch signal and other two used for providing heating current. Load resistance used in the sensor output circuit varies gas to gas. Generally, the output voltage is zero for fresh air and voltage level changes like detecting noise when gas exists. So, refereeing zero voltage at for fresh air means 0% gas existence, a proportional gas level calibrated for different percentage of gas existence [9].

D. Sensing of obstacle and position

The robot unit is equipped with obstacle avoiding and position sensing features for easy of operation. An ultrasonic sensor installed in front of the rover helps it to detect possible obstacle on its path and avoid them during a journey. A GPS receiver detects the position of the robot and updates it to the processor every second.

IV. ROBOT UNIT CIRCUITRY

A circuit diagram having each component of the robot sections has been depicted in the following diagram. Here the connections of the different sensors with the microcontroller and the power supply of the robot circuitry have been illustrated.

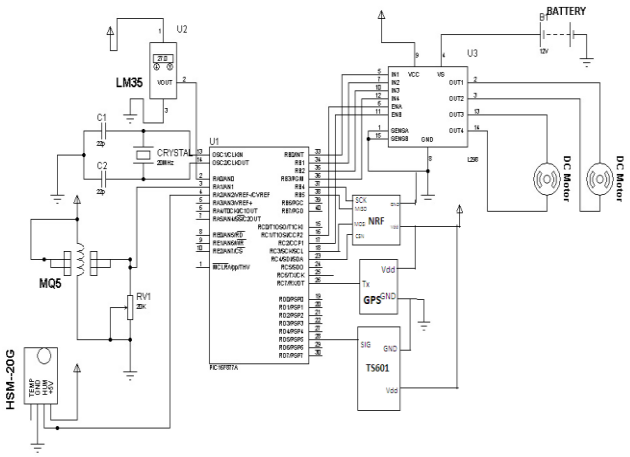


Fig. 4: Circuit diagram for robot unit

V. Firmware Description

The firmware development for the survey system can be split into two parts- developing firmware for the wireless control unit and another for the robot unit. The firmware for the control unit requests wireless connection establishment when started first. After establishing a successful connection between robot and control unit, the control unit firmware executes two different threads of operations. In one thread, it seeks for the remote data acquired by the robot on wireless communication terminal, receives them and displays them on the LCD. In the other thread, it takes user input from the operator (given using keypad), tags the input command and transmits them to the robot through wireless terminal.

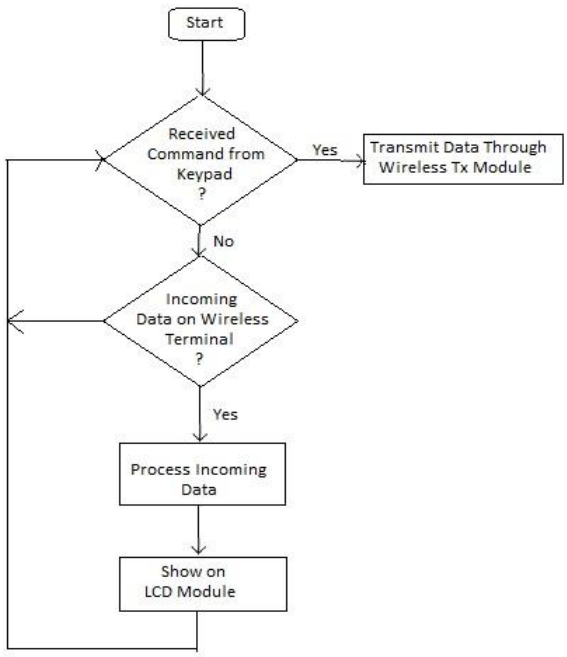


Fig. 5: Algorithom for control panel

The firmware for the robot unit is bit complicated then that of the control unit. When started, it first awaits for the control unit command to set the mode. After the mode is set by the operator (by wireless command), it starts executing the operations according to the mode. In manual mode, the firmware runs a program thread seeking wireless command continuously, and after a command is received, it executes the relevant operation according to the command. This commands involve the movement control of the robot, data acquisition command etc. If a command for changing the mode of the robot is received, it switches to the autonomous mode of operation, and takes the steering of the survey itself. From the ultrasonic sensor information, it detects obstacles in the path of the robot and controls the movement such that the robot can avoid them. After performing a random survey operation for specified time period, it leads the robot to come back to the starting point with the help of GPS receiver value. Different user defined functions are developed for interfacing with different peripheral modules interfaced to the microcontroller.

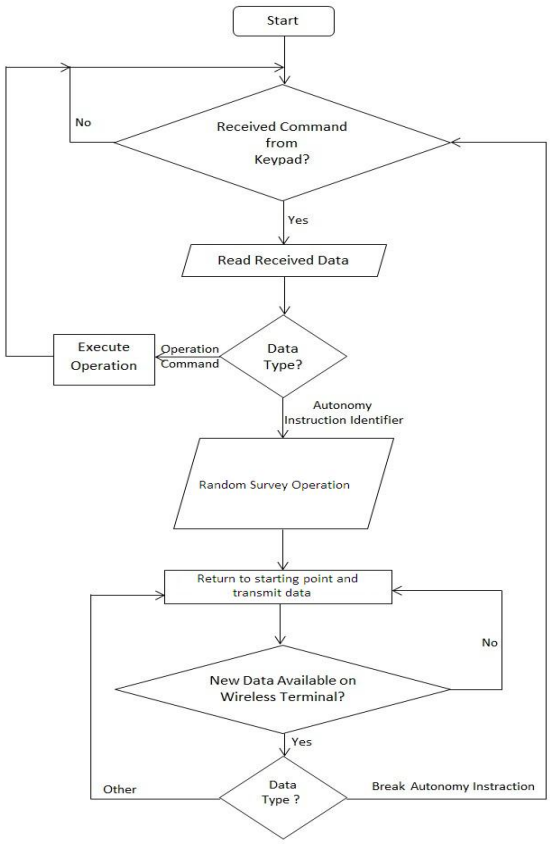


Fig. 6: Algorithom for Remote Survey Unit

VI. Performance Evaluation

The survey operation system depicted in this paper is a system with moderated range which covers area up to 100 meters of radius. Where most of the existing survey operation systems are of relative high range of operation, they often become expensive for uses in relatively less sophisticated environments, or in cases where a long range of operations are not always the key issue. The proposed system offers a relatively low cost solution in such cases. This also allows autonomy in random survey operation, which makes the system more comfortable in application field. Built in GPS device can be used to allow an autonomous returning to starting point, which provides extra security of the robot even in the worst case in radio dropout zones.

VII. CONCLUSION

The environmental survey assistant system of the paper has been designed and implemented for the convenience of the human being in any parlous places. It can operate both manual and autonomous mode which will give extra advantage in the survey operation in any condition. Though some systems existing for this kind of remote survey operations, but implementation of this system shows a cost effective and efficient survey system made up from commercially available components and implemented with intelligent control algorithms.

REFERENCES

[1] Theophilus Wellem, Bhudi Setiawan, "A Microcontroller- based Room Temperature Monitoring System", International Journal of Computer Applications (0975 – 8887), Volume 53– No.1, September 2012

[2] Mohd Adam, Mohd Razali, "Pc-Based Temperature Control", Universiti Malaysia Perlis, Malaysia, 19th April, 2007

[3] Xing Guo Quan, "Temperature Characteristics and Application of LM 35 Sensor", Biomedical Engineering Department of Xianmng College，Xianning Hubei-437100，China , Article id 1007—7510(2007)11—0049—02, 2007

[4] Gao Mei Zhen, "Principle and application of the LM 35 series temperature sensor", Hubei normal university，China, Article Id: 1671-1041(2005)01-0114-O 1, 2005

[5] JamesKramer, MatthiasScheutz, "Development environments for autonomous mobile robots: A survey", Auton Robot, Springer, December 1, 2006

[6] Matthew Dunbabin, Jonathan Roberts, Kane Usher and Peter Corke, "A New Robot for Environmental Monitoring on the Great Barrier Reef", CSIRO ICT Centre, Kenmore QLD 4069, Australia, 2007

[7] Youhua Chen, Gary Anderson, Samira Mahdi, Jarjees Khidir, Edmond Wilson, "A Proposed Robotic Air Quality Monitoring System", IEEE Transaction, 2012

[8] R.A. Hirschfeld, F. Aghazadeh, and R.C. Chapleskl, "Survey of Robot Safety in Industry", The International Journal of Human Factors in Manufacturing, Vol. 3 (4), 369-379 (1993)

[9] Gideon Kowadlo and R. Andrew Russell, "Robot Odor Localization: A Taxonomy and Survey", The International Journal of Robotics Research, August 4, 2008

Maze solving Algorithm for line following robot and derivation of linear path distance from nonlinear path

Shadman Sakib
Department of Naval Architecture & Marine Engineering
Bangladesh University of Engineering & Technology
Dhaka, Bangladesh
shadman.sakib_siam@yahoo.com

Anik Chowdhury, Shekh Tanvir Ahamed and Syed Imam Hasan
Department of Electrical & Electronic Engineering
Bangladesh University of Engineering & Technology
Dhaka, Bangladesh

***Abstract*— In this paper we have discussed a unique general algorithm for exploring and solving any kind of line maze with another simple one for simple mazes without loops or with loops having highest two branches none of which are inward. For the general algorithm, we need a method to map the whole maze, which is required if the maze is complex. The proposed maze mapping system is based on coordinate system and after mapping the whole maze as a graph in standard 'Adjacency-list representation' method, shortest path and shortest time path was extracted using Dijkstra's algorithm. In order to find the coordinates of the turning points and junctions, linear distances between the points are needed, for which wheel encoder was used. However, due to non-linear movement of robot, the directly measured distance from the encoder has some error and to remove this error an idea is built up which ended by deriving equations that give us almost exact linear distance between two points from the reading of wheel encoder of the robot moving in a non-linear path.**

***Keywords*—Maze solving, mapping, line following robot, wheel encoder, linear path distance.**

I. INTRODUCTION

In mobile robotics, maze solving problem is one of the most common problems and to solve this problem an autonomous robot is used. Mazes can be of different kinds, like- without any loops, having loops, grid system or without grid system. Numerous methods and algorithms have been developed having their own merits and techniques such as flood fill algorithm [3], breadth first-search [4], Prim's algorithm [5] etc. But they all need to know the position of the goal. So, they are not appropriate for solving mazes where the goal is unknown until discovering.

If there is no loop and we do not need to find the shortest path, then we do not need to map the whole maze for reaching the destination. This holds still true for the maze having 'small loop'. Here, by 'small loop' we mean loop with no inward branches and there can be highest two paths for getting in or out of the loop. And the loop will not contain the end-point.

In this work we have introduced a new method of maze solving for any kind of mazes. In order to solve a complex line maze, the main problem arises is to map the whole maze. Without mapping the whole maze, robot cannot find the shortest path. So, our main target is to map the whole maze and then find the shortest path. Our first algorithm is to solve simple mazes fulfilling the criteria mentioned before. Actually the first algorithm is an upgraded version of the most common and ancient maze solving method- 'Follow-the-right' or 'Follow-the-left'. The second algorithm is designed to solve and find the shortest path for any kind of maze. We have mainly focused on the mapping method using coordinate system and direction, by which we can save the whole maze as a graph in standard 'Adjacency-list representation' method [5]. Using this method the robot will get all adjacent points of each point and their coordinates. Then it will generate the desired graph of the maze in G = (V, E) format. Wheel encoder is used to determine the distance and the coordinates of the junctions and the turning points. However using wheel encoder value directly may bring some error because of non-linear movement of robot. Hence, we also discussed about the process of linearization of the distance that is computed from the wheel encoder reading.

II. 'NO-LOOP' OR 'SHORT LOOP' MAZE

In this algorithm, the robot is instructed to follow a preference of directions. There may be two sets of sequences of preference. They are- (1) 'Left-Front-Right-Down' and (2) 'Right-Front-Left-Down'. For example, we may mark directions as follow: (1) for right, (2) for up, (3) for left and (4) for down. Whenever a junction is encountered, the robot will always choose a path based on the preference sequence. Let, there are two paths coming out of one path, one goes to the left, another to the right. If the robot comes across it, it will first go to the left path (For sequence 1) to explore it. This junction and taken direction will be saved in the memory. If after following the path, it comes to a dead end it will turn back and come to the previous position. After reaching that position, previously saved value in the data structure will be extracted and the present preferred direction number will be added to that value. This new value will be saved again in the same place of the data structure. Same procedure will be followed continuously until it finds the 'end-point'. So, the robot will have a record of

number of junctions it came across and a corresponding array to record the taken direction. Moving like this, one time the robot will reach the 'end-point' and the recursively updated direction array will generate an error free simple path from starting point to finishing point. Here, is an example showing how it works−

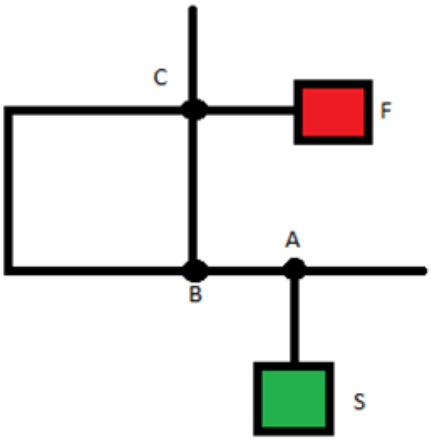


Fig. 1 An example of generalized algorithm for 'no-loop' or 'short loop' maze (showing only the points of interest)

For example we have a maze in the fig.1 (showing only the points of interest). The robot starts roaming from point 'S' and stops at point 'F'. We need an integer to keep record of present junction and an array to save the movement at each junction it came across. The preferred sequence of the robot is 'Right-Front-Left-Down'. Initially, present junction number and each elements of the array are 0. Now when the robot comes to junction 'A', at first it increases the present junction number by 1. Here it has a path at right (1) so it takes that path, add 1 to the current value of the 1st element of the array (As the present junction is 1), which is 0 and save it in that place. So, the value of the 1st element of the array will be 1. Following the line it will encounter a dead end and turn back. As it is turning back to the previous path, present junction will be decreased by 1 to 0. When it comes to the junction 'A' again, First, it will again increase present junction to 1. Now, it has two paths- one in the front (2) and another to the left (3). According to the preference, it will take the straight path and preferred direction number (2) will be added to the previously saved value in the 1st element of the array. Therefore its current value of that element will be 3. Now when it comes to junction 'B', present junction will be 2. Here it will find a path in the right (1) so it will take that path and add 1 to the value of the 2nd element of the array. So, value of the 2nd element of the array will be 1. When it comes to the junction 'C' same procedure will happen and eventually it will come to the 'end-point'. So after the exploration, the array will look like this – [3 1 1] (Indexing from 1). Hence, starting from the beginning when it comes to junction 'A' it will take the direction indicated in the first element of the array which is 3, thus left (3) and when it comes to the junction 'B' it will go right ('1' is for right) and at the junction 'C' it will again take the right ('1', for right) path and will come to the end-point, which is the shortest path of the explored paths by the robot.

III. Algorithm for Solving Any Kind of Maze

This solving technique is developed for the case where we do not have any grid system, there is no restriction about the loops and we have to find out the exact shortest path. It is said before that our main concern here is to map the whole maze and store it as a graph in standard 'Adjacency-list representation' method. In this method, we must know the adjacent points of each point. Then we can solve the graph for shortest path using Dijkstra's algorithm [2]. There are two types of points we need to know about- turning points and junction points.

In this algorithm, there will be four arrays. Among them one is two dimensional. This two dimensional array is used to store the coordinates (x, y) of the points. Rest arrays are vectors. They are –

A. *Point array* - saves each point by a distinct number or character in the serial they are explored.

B. *Type array* - identifies the type of the corresponding point of *Point array*.

C. *Explored array* - keeps record how many times a junction point is differently explored by the robot.

Elements of the *Type array* is one less than the number of ways a point directs towards. It actually indicates how many times we need to go through that point (each time choosing a different path) to explore all adjacent points of the respective point. Again, the number of paths we have already explored will be saved in the *Explored array*. As long as for each point the values of their corresponding element of *Explored array* and *Type array* are not same; we have to continue scouting to map the maze because there are still some ways or even points to be explored.

As mentioned earlier, there will be four arrays and two variables i.e. *Total_points* and *Direction*. At the beginning of the exploration the robot encounters the starting point. We name this point as '0'. So the first element of *Point array* will be 0 (*Point* [0] = 0) and the variable *Total_points* will be 1. The robot will also save the coordinate of that point in *Coordinate array* i.e. *Coordinate* [0] [0] = 0, *Coordinate* [0] [1] = 0. Here the first indexing element of the array represents the point number and the second one represents the x-coordinate or y-coordinate. The robot will save a point's x and y coordinates in its corresponding position in this 2D array. By doing this, it will get the point number and its corresponding coordinates. Now, it will take a direction and again the value of the variable *Direction* will change. In this algorithm, as we are to map the whole maze, absolute direction is used i.e. North-South-East-West. Say if it takes an eastern direction then the value of *Direction* will be (1), for north it will be (2), (3) will be for west and (4) for south. Values can be different, but increment must be in a cyclic order. Initial direction will be regarded as north. After taking a direction it will start to read the encoder value and follow the line until a junction or turning point is reached. After reaching a junction or turning point it will stop there and calculate the distance and identify coordinate. Now two situations may arise whether it can be a new point or an old one. In order to find whether it is a new point or an old one, we cross check the coordinates of that point with the elements

of *Coordinate array* within a predefined tolerance value. If the point is a new one, then the robot will save it in the *Point array* with a new name, increases the *Total_Points* value, save type value in the *Type array*, save the ways it has been explored in the corresponding element for that point in the *Explored array* and also save coordinates of that point in *Coordinate array*. Then it will again begin to explore and take a direction from the sequence of preference. For exploration it will do the same thing again and again saving all kinds of values for each point i.e. *Point* number, *Type*, *Explored* number, *Coordinate* and *Direction*. If the point is not a new one then it will save the point again in the *Point array* with its previous name, find the first position of that point in the *Point array* and increase the value of its corresponding element of *Explored array* if last point does not exist before adjacently to this point in the *Point array*. Value of the element of *Explored array* for any corresponding point is equal to one less than the number of its explored total different adjacent points. If the *Explored* value is less than the *Type* value then it will take the unexplored path and explore it. If the *Explored* value of that point is equal to the *Type* value then it will search the *Point array* and find those points of which *Explored* value is less than *Type* value. If there is any point like this, then it will go to explore it by using the *Point array*, *Direction* and *Coordinate array*. If there is no point then the exploration is finished. By doing this recursively it will explore the whole maze and get all the junctions and turning points, their coordinates and their adjacent points, which are necessary to build the graph.

Case study: Here is a maze for demonstrating the exploring technique of the robot. (fig. 2) At first the robot starts from the starting zone, gives it a name in *Point array*('S'). (It is more convenient if we use integer to number the points serially like 1, 2, 3… instead of A, B, C… But here, for easy representation, we have used Characters) and identify the coordinate of this point as (0, 0) and save it in the *Coordinate array*. As it is a dead end, the *Type* value will be 0 and the *Explored* value will be 1. Now the robot is in the forward direction which will be considered as north. So the *Direction* value will be 2 as we assumed earlier and the *Total_Points* value will be 1. Now it will go to forward direction and start taking the encoder value. When it will reach a point, it will check the coordinates of that point. As the robot came from (0, 0) point to north side and say total distance is 10 units, that means it moved along the y-axis in positive direction. So the coordinate will be (0, 10). It does not match to the coordinates of any of the previously visited points. So it is a new point. As it is a four way junction, the value of *Type* will be 3 and the *Explored* value will be 1. Now suppose it prefers the eastern path. So, it will take a right turn and the value of *Direction* will be updated now. It will again begin to explore and after reaching the new point, it will identify its coordinate 'E' (14,10), of which *Type* value will be 2. Now it will again take the northern direction and find new point 'D' (14, 13). Similarly, it will explore the 'G' (14, 16) point. Here it will search for the points having smaller *Explored* number than their *Type* number and come back to 'D', as it is one of them and the closest one. Now, as this is an old point, it will find out the explored adjacent points of this point from the array which are 'G' and 'E'. Analyzing their coordinates it will know that they are in north and east side and the only available new path

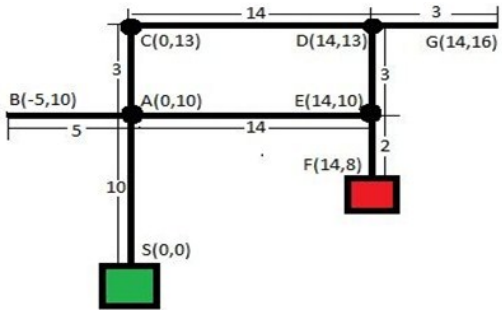


Fig. 2 An example of maze for proposed maze mapping algorithm for any kind of maze (with showing all the points and distance between points)

is the western one. So, it will go through that path. D's *Explored* number will be now 2; equal to the *Type* number. After a while, the robot will again reach 'A' point passing 'C'. Now, as it came here through a different path, *Explored* number will be increased. After analyzing the adjacent points' coordinates, western path will be chosen. *Explored* number will be again increased. So, it will be now 3; equal to the *Type* number. Eventually it will explore point 'B' which is a dead end. So, here, it will search for the points having smaller *Explored* number than their *Type* number. Only one such point exists – 'E'. So, it will go to the 'E' point and take the southern path, which leads to the dead-end end-point 'F'. Now, there is no points available having smaller *Explored* number than *Type* number. So, the exploration is finished.

IV. Generating Graph of Standard Format and Finding the Shortest Path

The last stage of our mapping algorithm is making an adjacency-list for each point from the obtained data. We can easily obtain this from the *Point array*. To do so, we will take one particular point from that array and mark each position of that point in the *Point array*. Then we will find out all points which are in the immediate next or previous position to that particular point. These are the adjacent points to that particular point. Doing this for each point will give us the complete 'Adjacency-list representation' of the graphed maze in G = (V, E) format. To determine the node to node distance, we will also need the coordinates of each point which we have already stored in the Coordinate array. After that, using A* or Dijkstra's or any other algorithm we can extract the shortest path between nodes. The following figure of the data structure will help to realize this process –

Point	S	A	E	D	G	D	C	A	B	A	E	F
Type	0	3	2	2	0	2	1	3	0	3	2	0
Explored	1	3	2	2	1	2	1	3	1	4	2	1
CoOr_X	0	0	14	14	14	14	0	0	-5	0	14	14
CoOr_y	0	10	10	13	16	13	13	10	10	10	10	18

So, A ⟶ S, E, C, B

Fig. 3 Data structure of generating the whole maze as G = (V, E) format from the explored points of the maze with corresponding coordinates. (Coordinates are determined by the corresponding point to point distance obtained by using wheel encoder and the equations of distance which is derived in the later part of the paper with taking account some minor error. Generally the error is ±3%)

V. Obtaining Precise Distance from Wheel Encoder

In grid less mazes, a robot must know its current position and calculate distance between nodes. Not only in maze solving, but also in many other fields, it is important to measure distance precisely using wheel encoder. However, while measuring the length of a straight line, wheel encoder cannot give us the exact value directly. This is due to the fact that a line following robot does not always go along the straight line. The path followed by the robot is rather zigzag. There are two reasons behind this behavior — (1) Imperfect alignment of robot with the line which causes linear deviation from the line. (2) Difference between the RPMs (Revolutions per minute) of the driving motors which causes circular deviation from the line.

Here, we inspected the reasons behind the curved path of the robot and derived equations to extract the linear distance between two points from the computed zigzag path distance.

A. Initial Alignment of Robot

A robot does not remain in perfect alignment with the line when it starts moving. So, it goes in a straight line (AJ in the fig. 4) making an acute angle with the line to be followed (AB in the fig. 4), if the speeds of the driving motors are same. Line detecting sensors detect the deviation when a minimum distance ('h' in fig. 4) is created between the middle sensor (midpoint of the robot's front side) and the line.

Then, the robot makes a right turn making an angle of θ with the path. The robot will go through this new line until again it creates h distance from the line in opposite distance. As the distance is same, the rotation will be same as the previous one (θ). Because it is a conditional rotation, not time based. This will continue. So, the path followed by the robot is like the fig. 4.

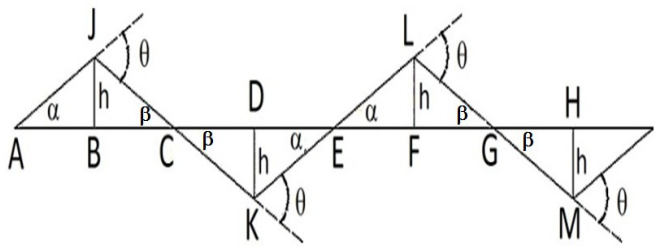


Fig.4 While following a linear path, robot deviates from the straight line if it is not fully aligned with the line. So, the robot goes to left side with an angle of α with the line to be followed until the sensors detects the deviation and again turns to the right side with an angle of θ with the line and the path the robot is moving looks like zigzag path.

In fig. 4, as rotation is same in points J and K, AJ and KL are parallel. Now, in ΔABJ, AB = AJ × $cos\alpha$. Similarly, BC = CJ × $cos\beta$, CD = CK × $cos\beta$.

Adding them all together, we get-

AB + BC + CD + DE + EF + FG + GH = (AJ + KE + EL) × $cos\alpha$ + (JC + CK + LG + GM) × $cos\beta$.

So, linear distance ≈ (C_β × Total wheel distance after right turn) + (C_α × Total wheel distance after left turn)

Where, $C_\alpha = cos\alpha$ and $C_\beta = cos\beta$.

However, measuring distance so frequently (after each rotational move) generates more error. To mitigate this problem, we have used one constant (C) instead of C_α and C_β. So, we can write-

AB + BC + CD + DE + EF + FG + GH = (AJ + JC + CK + KE + EL + LG + GM) × C

However, when a robot changes direction with rotational move, the wheels crosses some distance in a nonlinear, circular path. In fig. 5, left wheel of the robot goes through JD (straight line) until deviation reaches h, then it goes in an arc shaped path DE. Then again goes in a straight path EH. This goes beyond our assumption in fig. 4.

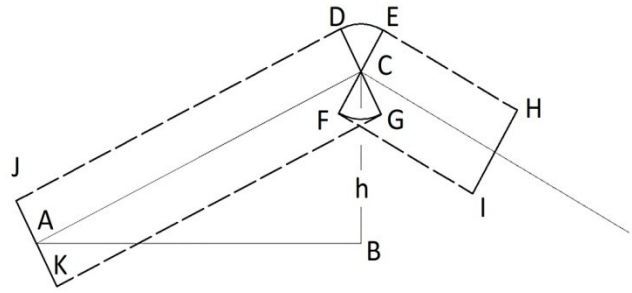


Fig. 5 When changing direction with rotational move, the wheel crosses some distance in nonlinear circular path.

So, we have to change our equation for extracting linear distance—

$$D = [WL_{Total} - F_{LC} \times (N_R + N_L) - K \times N_L] \times C_L + F_{LL} \times N_R \quad \text{(For left wheel).} \quad (1)$$

Or,

$$D = [WL_{Total} - F_{RC} \times (N_R + N_L) - K \times N_R] \times C_R + F_{RL} \times N_L \quad \text{(For right wheel).} \quad (2)$$

Where, F_{RC} and F_{RL} are respectively the circular and linear distance that the right wheel crosses while rotating.

F_{LC} and F_{LL} are respectively circular and linear distance that the right wheel crosses while rotating.

N_R and N_L are number of total right and left turns respectively.

C_L and C_R is the value of C for left and right wheel respectively.

WL_{Total} and WR_{Total} are total distance crossed by left wheel and right wheel respectively and K is inner rotational constant.

B. Speed difference of the driving motor

Having motors of exact same RPM is unlikely. Even after using PWM (Pulse Width Modulation), there remains a little difference in RPM between two motors. This speed difference leads to deviation from straight line and causes curved path.

If we let the robot go freely where, V_R (RPM of right motor)> V_L (RPM of left motor), the robot will follow a circle.

If the radius of the circle travelled by the left wheel is R' and R'''' for the right wheel then as in same time the wheels travelled the perimeter of these circles, $V_R \times T = 2\pi R''''$ and $V_L \times T = 2\pi R'''$.

So, $\frac{V_R \times T}{V_L \times T} = \frac{2\pi R''''}{2\pi R'''}$ or, speed ratio $= \frac{R''}{R'} = 1 + \frac{B}{R'}$. (3)

Where, B = Distance between two wheels.

So, the robot will always maintain same radius while moving in a circular or arc shaped path. This radius will increase if the difference between the RPM of the motors decreases. It is assumed that the RPM difference is very small. However, the path line of the robot among successive two rotational moves is not linear, rather circular. So, the path line of the robot will be similar to fig. 6. This means, the distance we get from the wheel encoder, is of an arc length rather than a straight line.

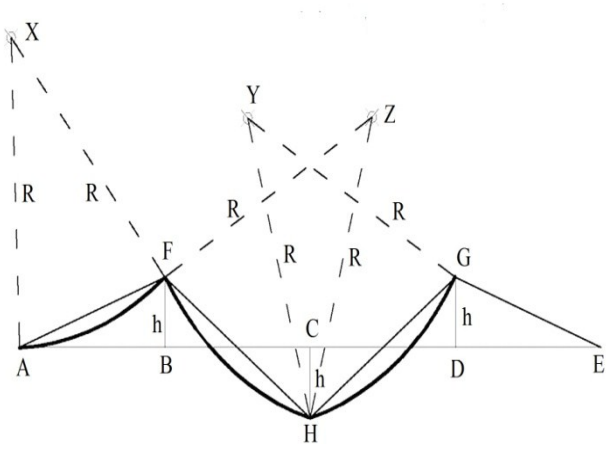


Fig. 6 As the RPMs of two motors are not identical the robot will follow the line not in a straight line rather in arc shaped line as the speed ratio (from equation 3) is not unity

If the equation of the circle be $x^2 + y^2 = R^2$, considering the first portion of fig. 6. Arc length,

$$ds = \int \sqrt{(dx)^2 + (dy)^2} \text{ or } s = \int_x^{x+h} \sqrt{(1 + \left(\frac{dx}{dy}\right)^2)}\, dy$$

Again, differentiating the equation of circle gives - $\frac{dx}{dy} = -\frac{y}{x}$. From this value, we get $s = \int_{x_1}^{x_2} \sqrt{(1 + \frac{x^2}{y^2})}\, dx = R\left[sin^{-1}(\frac{x}{R})\right]_{x_1}^{x_2} = R\left\{sin^{-1}(\frac{\Delta x}{R})\right\}$ [$\Delta x = x_2 - x_1$] or, $\Delta x = R\, sin(\frac{s}{R})$.

This equation can extract length of the straight line from the length of the curved path the robot follows. However it can only be used when the path is a continuous arc. So, we have to read wheel encoder before each rotation and calculate the distance from last rotation point, which is not an efficient way and yields to very high error rate. So, we need to find out another way so that we can use the total travelled distance to find out the length of the straight line at once. From fig. 4 and fig. 7, we can see that the distance, a robot travels along the y-axis while moving in a continuous arc is of height either h or 2h unit. The arc length of an arc of height h can be calculated from this formula: $S_h = \int_y^{y+h} \sqrt{(1 + \frac{y^2}{x^2})}\, dy = R\, sin^{-1}(\frac{h}{R})$.

For arcs having ordinate length of 2h, the equation of arc length becomes: $S_{2h} = R\, sin^{-1}(\frac{2h}{R})$ (4)

Now, when robot goes in a zigzag way we considered the $[WL_{Total} - F_{LC} \times (N_R + N_L) - K \times N_L]$ term as a linear distance, which is not true if the RPMs of both motors are not same.

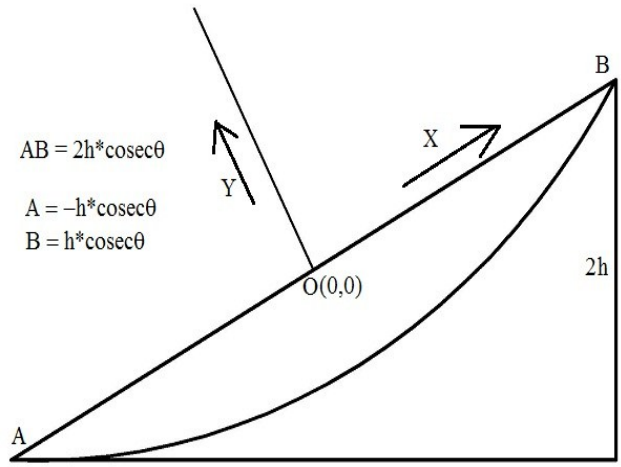


Fig. 7 The distance the robot travels along the y-axis while moving in a continuous arc is of height either h or 2h. Calculating the length of the arc path and corresponding x-axis direction mitigates the rate of error.

This is a sum of several arc lengths. Among them, (NR + NL − 1) are of 2h height, 1 of h and another is not fixed. Now, we can modify that equation to get a more correct one. From fig.7 we can say, $s_{2h} = \int_{-hcosec\theta}^{hcosec\theta} \sqrt{(1 + \frac{x^2}{y^2})}\, dx = 2R\, sin^{-1}(\frac{X_{2h}}{2R})$.

So, $X_{2h} = 2R\, sin(\frac{S_{2h}}{2R})$ *and* $X_h = R\, sin(\frac{S_h}{R})$ (5)

Arcs having 2h and h heights have a total length of $(N_R + N_L - 1) \times S_{2h} + S_h$.

So, arc length of last stage, $S_D = [WL_{Total} - F_{LC} \times (N_R + N_L) - K \times N_L] - (N_R + N_L - 1) \times S_{2h} - S_h$.

This is an arc which can be of various lengths. So,

$D_D = R \times sin(\frac{S_D}{R})$. (6)

Now, we can find the equations for linear distance measurement-

$D = [(N_R + N_L - 1) \times X_{2h} + X_h + D_D] \times C_R + F_{RL} \times N_L$ (For right wheel). (7)

$D = [(N_R + N_L - 1) \times X_{2h} + X_h + D_D] \times C_L + F_{LL} \times N_R$ (For left wheel). (8)

These equations are the final equations which lead us to the exact length of a straight line from wheel encoder reading of a line follower robot. This equation also shows that if we know the minimum deviation for turn (h) and other constants (C_R , F_{LL} ... *etc*) with the total turn, we can predict the path

distance to a fair approximation without even knowing the travelled distance. That means, for rough calculation of distance, wheel encoder is not needed at all.

VI. RESULTS

Our proposed mapping technique along with Dijkstra's algorithm solves mazes of any shape. However mazes with large number of nodes need IC capable of high computational power. Again implementing the proposed equations we derived following data was found –

TABLE I. SAMPLE MEASURED PATH DISTANCE

Type of Length	Test Number		
	1st test	*2nd test*	*3rd test*
Actual length (cm)	10	14	8
Length from wheel encoder (cm)	10.23	14.14	7.90
Length from formula (cm)	10.02	14.04	7.99

If we include Kalman filter with this, error will be more less. So, we will get an very efficient system. But from fig. 8 we can see that the signal from wheel encoder is pretty much steady. So, filter is not necessary for the encoders. But it may be applied to the line following portion.

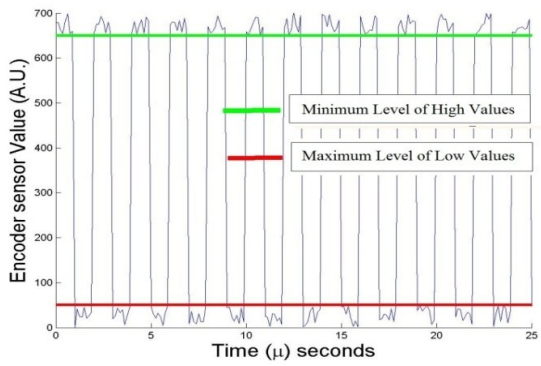


Fig. 8 The ADC (Analog to Digital Converter) value of wheel encoder suggests that further normalization is not needed for encoder value as the rate of error is negligible.

In many maze solving algorithms, it is important to know the position of the goal. So, they are not appropriate for solving line mazes where the goal is unknown until discovering. The proposed algorithm mitigates this limitation.

VII. CONCLUSION

Different types of maze solving algorithm are often needed for different mazes. However there is no perfect general algorithm for that purpose. The algorithm discussed in this paper will give us a general method applicable for almost any kind of maze. The algorithm can also be used in sectors beyond line following. After some modifications, it can be used for autonomous road mapping, cave mapping, exploring and many other purposes. In this cases low cost GPS devices can be used instead of wheel encoder for tracking coordinates. Again we often need to determine the linear path distance using wheel encoder of a robot which brings error because of the nonlinear motion of the robot. Using the equations derived here, we can find it accurately.

REFERENCES

[1] Rahnama Behnam, Ozdemir Makbule Canan, Kiran, Yunus, Elci, Atilla, "Design and Implementation of a Novel Weighted Shortest Path Algorithm for Maze Solving Robots", Proceedings of the IEEE 37th Annual Computer Software and Applications Conference Workshops (COMPSACW), pp. 328 - 332 (2013).

[2] Huijuan Wang, Yuan Yu and Yuan, Q. "Application of Dijkstra Algorithm in robot path planning", Proceedings of the Second International Conference on Mechanical Automation and Control Engineering (MACE), pp. 1067-1069 (2011).

[3] Nosal, E. "Flood-fill Algorithms Used for Passive Acoustic Detection and Tracking", Proceedings of the IEEE International Conference on Digital Object Identifier, pp. 1-5 (2008).

[4] Knuth, Donald E. "The Art of Computer Programming", Vol.1, 3rd Edition, Boston: Addision-Wesley (1997).

[5] Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest and Clifford Stein "Introduction to Algorithms", 3rd Edition, Chapter VI – Graph Algorithms, pp. 587-768, The MIT Press (2009).

[6] Sadik, A.M.J., Dhali, M.A., Farid H.M.A.B., Rashid, T.U. "A Comprehensive and Comparative Study of Maze-Solving Techniques by Implementing Graph Theory", Proceedings of the IEEE International Conference on Artificial Intelligence and Computational Intelligence (AICI), Vol.1, pp.52-56 (2010).

[7] Hongshe Dang, Jinguo Song and Qin Guo "An efficient algorithm for robot maze-solving", Proceedings of the 2nd International Intelligent Conference on Human-Machine Systems and Cybernetics (IHMSC), Vol.2, pp.79-82 (2010).

[8] Mishra, S. and Bande,P. "Maze solving algorithm for micromouse", Proceedings of the IEEE International Conference on Signal Image Technology and Internet Based Systems , pp. 86-93 (2008).

[9] Cai, J., et al. "An Algorithm of Micromouse Maze Solving", Proceedings of the 10th IEEE International Conference on Computer and Information Technology, pp. 1995-2000 (2010).

[10] Ya Juan Zhang, Zi Li Zhang, Yong Deng "An improved maze solving algorithm based on an amoeboid organism", Proceedings of Control and Decision Conference (CCDC), pp.1440-1443 (2011).

[11] Xihua Li, Xiang Jia, Xudan Xu, Jin Xiao, Haiyue Li "An Improved Algorithm of Exploring Process in Micromouse Competition", Proceedings of the IEEE International Conference on Intelligent Computing and Intelligent Systems(ICIS) Vol.2, pp. 324-328 (2010).

[12] Andrew M. Rockett "Arc Length, Area and Arcsine Function", Proceedings of the Mathematics Magazine, Vol. 56 , Number 2, pp. 104-110 (1983).

[13] Zhanna V. Zatuchna, Anthony Bagnall "Learning Mazes with Aliasting States: An LCS Algorithm With Associative Perception", Vol.17,1, pp.28-57 (2009).

[14] Osmankovic D., Konjicija S. "Implementation of Q – Learning algorithm for solving maze problem", Proceedings of the 34th International Convention MIPRO, pp.1619-1622(2011).

[15] Zhuang Cai, Lu Ye, Ang Yang "FloodFill Maze Solving with Expected Toll of Penetrating Unknown Walls for Micromouse", Proceedings of High Performance Computing and Communication & IEEE 9th International Conference on Embedded System and Software(HPCC-ICESS), pp. 1428-1433 (2012).

[16] Jiang Hualong, Wang Hongqi, Tian Yonghong "Design and realization of a maze robot", Proceedings of the International Conference on Consumer Electronics, Communications and Networks (CECNet), pp. 201-204 (2011).

[17] Kazerouni, B.H., Moradi, M.B.and Kazerouni, P.H. "Variable priorities in maze-solving algorithms for robot's movement", Proceedings of the IEEE International Conference on Industrial Informatics, pp. 181 – 186 (2003).

EEG Signal Classification using Frequency Band Analysis towards Epileptic Seizure Prediction

Mohammad Zavid Parvez
School of Computing and Mathematics
Charles Sturt University
Bathurst, Australia
Email: mparvez@csu.edu.au

Manoranjan Paul
School of Computing and Mathematics
Charles Sturt University
Bathurst, Australia
Email: mpaul@csu.edu.au

Abstract— **Epilepsy is one of the most common and diverse set of chronic neurological disorders characterized by an abnormal excessive or synchronous neuronal activity in the brain that is termed "seizure", affecting about 50 million individuals worldwide. *Electroencephalogram* (EEG) signal processing technique plays a significant role in detection and prediction of epileptic seizure. Recently, many research works have been devoted to detect/predict of epileptic seizure based on analysis of EEG signals. Even though remarkable works have been conducted on seizure detection/prediction, experimental results are not mature enough in terms of sensitivity, specificity, and accuracy. In this paper we present a new approach for seizure detection to analysis preictal (before seizure onset) and interictal (period between seizures) EEG signals by extracting different features from gamma frequency band by decomposing the signals using discrete wavelet transformation. Note that the detection of preictal and interictal EEG signals leads to predict the epileptic seizure. Experimental results demonstrate that the propose method outperforms the state-of-the-art method in terms of sensitivity, specificity and accuracy to classify seizure by analyzing EEG signals to the benchmark dataset in different brain locations.**

Index Terms—EEG, DWT, LS-SVM, Seizure, and Epilepsy.

I. INTRODUCTION

Seizure is sudden, intensive, unexplained, and unknown neurological activity in the brain. Epilepsy is a medical condition having spontaneous recurrent seizures [1]. Epilepsy takes place when many neurons are excited in the same time due to brain injury or chemical imbalance in the brain that is characterized predominantly by unpredictable interruptions of normal brain function. Epilepsy puts patients in many injuries including high risk of fractures, submersion, burns, motor vehicle accidents and even death. Epilepsy affects approximately 1% of the world's population [1]. It is highly possible to prevent epilepsy if we can reliably predict the pre-syndrome of the epilepsy with high sensitivity and specificity.

Human brain is an electro-chemical machine, because neurons use chemical reaction to generate electricity. *Electroencephalogram* (EEG) is the well acceptable and most common diagnosis process for epilepsy and can detect abnormalities in the brain's electrical activity. It is highly possible to predict/detect seizure onset through analysis of the EEG signals. The prediction of epilepsy is clinically apparent in preictal (i.e., before seizure onset) stage of epilepsy. Timely and properly prediction of epilepsy can avoid fractures, submersion injuries, burns, and vehicle accidents, etc. of a patient. Therefore, it is important to prevent the epilepsy by analyzing EEG signals before the actual seizure onset.

Many methods have been proposed to predict epileptic seizure by classifying preictal and interictal (i.e., period between seizures) EEG signals. They have employed univariate techniques [3], eigenspectra of space delay correlation and covariance matrices [4], Hilbert-Huang transform [5], and autoregressive modeling and least-squares parameter estimator [6]. According to the classification results of preictal and interictal EEG signals, Chisci *et al's*. [6] technique carries good sensitivity compared to above mentioned techniques [4][5] using dataset [7]. However, Rasekhi *et al*. [3] got 79.3% sensitivity for their own dataset. The above mentioned techniques [4]-[6] have good sensitivity (i.e., preictal detection), however, the specificity (i.e., interictal detection) and accuracy (both preictal and interictal detection) are not expected level.

Recently, many techniques have been proposed for correct seizure detection. Among the existing techniques, Bajaj *et al.*'s [8] technique is newest and preeminent in term of performance and it can also measure non-linear dynamics of the EEG signals in proper way using features i.e. amplitude and frequency modulation bandwidth through *empirical mode decomposition* (EMD) technique. We have observed that Bajaj *et al.*'s [8] technique exhibits almost perfect classification accuracy for the small dataset [9] of seizure (i.e., ictal) and non-seizure EEG signals. However, their performance is below the expected results using the large dataset [7] for preictal and interictal EEG signals classification.

Bajaj *et al.*'s [8] have decomposed the EEG signals into different *intrinsic mode functions* (IMFs), where they did not explicitly decompose the signal into different frequencies. However, a number of literature including [10] has decomposed the signal into different frequency bands and then extracted different features from different bands to get band specific features for seizure detection/prediction. It is quite understandable that all bands are not related to seizure. Thus, features from some bands are highly related to the seizure compared to that of other bands.

Moreover, with improvement of technologies and

increasing number of quality channels, it is essential to realize patterns that are potentially involve in all the EEG channels across a range of temporal scales [4]. Temporal correlation of the EEG signals carries good sensitivity for preictal signal detection which leads towards epileptic seizures prediction.

In this paper, we present a novel approach for feature extraction and classification through *discrete wavelet transformation* (DWT) to decompose a signal into different frequency bands such as gamma (30-60Hz), beta (13-30Hz), alpha (8-12Hz), theta (4-8Hz), and delta (0-4Hz). Among the frequency bands, the gamma band is related to seizure as it carries more information than the other bands. This gamma band is used for feature extraction to exploit temporal correlation of the signals. These features are used as an input to *least square support vector machine* (LS-SVM) for classification of preictal and interictal EEG signals which lead toward seizure prediction. The experimental results show that newly extracted features provide better classification accuracy compared to the existing state-of-the-art method [8] for the classification of preictal and interictal EEG signals using benchmark large size datasets [7] in different brain locations.

II. Proposed Method

In this paper, we proposed a classification technique using two features namely frequency and amplitude from the gamma band signals through DWT decomposition technique considering temporal correlation and then classify preictal and interictal signals using LS-SVM. We have used a standard dataset [7] for epilepsy detection and prediction.

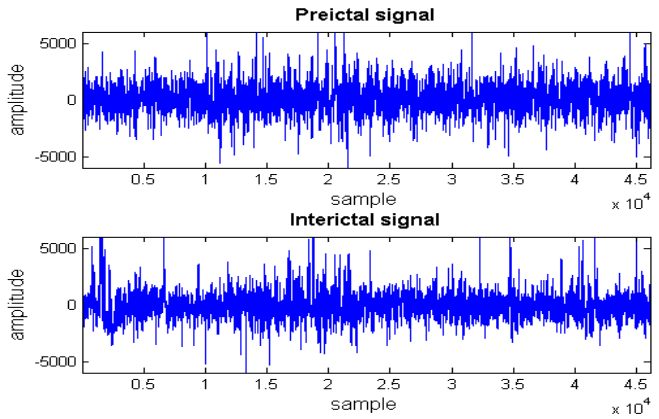


Fig. 1. Preictal (patient one, 7th datablock, channel one) and interictal (patient one, 95th datablock, channel one) EEG signals from Frontal lobe [7].

A. Dataset

The dataset used in the paper was recorded at Epilepsy centre of the University Hospital of Freiburg, Germany [7]. The database contains invasive EEG recordings of 21 patients suffering from medically intractable focal epilepsy. The data obtained by Neurofile NT digital video EEG system with 128 channels, 256 Hz sampling rate, and 16 bit analogue-to-digital converter. According to the dataset [7], epileptic EEG signals can be classified into (i) ictal (seizure period), (ii) preictal (before seizure onset), and (iii) interictal (period between seizures). Duration of a typical seizure is varied from few seconds to 2 minutes. However, according to the dataset, duration of an ictal signal is around 60 minutes containing seizure and at least 50 minutes of preictal signal preceding each seizure. The median time period between the last seizure and the interictal signal is 5 hours and 18 minutes, and the median time period between the interictal signal and the first following seizure is 9 hours and 36 minutes [4]. As suggested in the dataset, we have treated preictal as 3-minute prior on seizure onset. Three minutes from each interictal signal was also taken from the dataset to provide a comparable sized sample. Three minutes are taken from the 7th minutes of an interictal signal. In this experiment, we have used EEG signals from 12 patients where 6 patients for Frontal lobe signals and 6-patients for Temporal lobe signals. Sample signal is shown in Fig. 1.

B. Feature Extraction

DWT works based on multi-level where each EEG signal is decomposed into several scales and each scale provides a particular coarseness of the signal. Each phase of decomposition of a signal is composed of two down samplers by 2 and two digital filters. In each phase, high pass filter which serves as the discrete mother wavelet, and low pass filter which acts as the mirror version of the corresponding. The down-sampled outputs of the first low-pass and high-pass filters supply the approximation *A1* and detail *D1*, respectively. The first approximation *A1* is further decomposed and the procedure is continued. Adeli *et al.* mentioned that the Daubechies order 4-wavelet (db4) is most suitable for EEG signals analysis [13]. In our experiment, we have used Daubechies wavelets db8 on the right to extract gamma band and then extract features from gamma band to classify preictal and interictal EEG signals.

The wavelet transform can better localize the signal components in time frequency space. The Morlet-Grossmann defined the continuous wavelet transform for one dimensional *x(t)* signal is as follows [14]:

$$W(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t)\psi^*(\frac{t-b}{a})dt \qquad (1)$$

where (.)* denoted the complex conjugate, $\psi(t)$ is the analysing wavelet, a is the scalar parameter, and b is the position parameter.

The most common method is to separate continuous wavelet to discrete wavelet by scale factor of binary scales. Let $a = 2^j$ and the equation is as follows [15]:

$$W(j,b) = \frac{1}{\sqrt{2^j}} \int_{-\infty}^{\infty} x(t)\psi^*(\frac{t-b}{2^j})dt \qquad (2)$$

The first split of signal *x(t)* is *A1* and *D1*. Then further *A1* is split into *A2* and *D2*. This procedure continues until specified level is reached. For example, A_i and D_i (i = 1, 2,.., n) are the low frequency component and high frequency component of the ith level respectively. The 4-level DWT is

shown in Fig. 2 and its output signals are shown in Fig. 3.

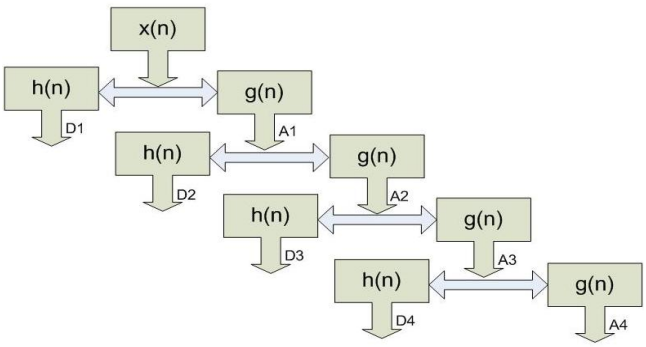


Fig. 2. EEG signals decomposition through 4-level DWT.

To extract the features using DWT, we calculate gamma band from three minute EEG signals and then reshape this gamma band into a two dimensional feature vectors to exploit temporal correlation of signals by dividing the signal into 15 seconds epoch (i.e., total 12 epochs for a signal). For gamma band of preictal, the values of frequency distribution is low and amplitude distribution is high whereas the characteristics of interictal signal are reverse as shown in Fig. 4. Thus, we consider average of up to one-quarter minimum frequency and one-quarter maximum amplitude as features. The average values of frequency and amplitude are used as input of the LS-SVM classifier for preictal and interictal classification.

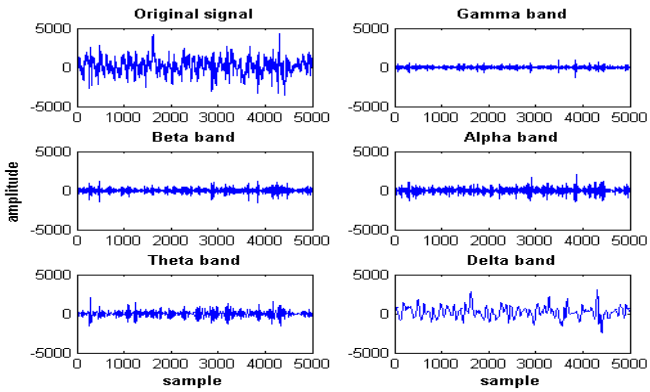


Fig. 3. Different frequency band after applying DWT decomposed technique in preictal signal of Frontal lobe signal (patient one, 7th datablock, and channel one).

C. Classification

The goal is to find patients states, preictal (Class 1) and interictal (Class 2) using the machine learning approaches through cross-validation evaluation. The challenge is to find the mapping that generalized from training sets and unseen test sets. For the cross-validation, data are partitioned into training set and test set. To classify the preictal and interictal signals, we have extracted features from gamma band using DWT. For classification, we have used LS-SVM [11] classifier as the LS-SVM is one of the best classifiers in the EEG signals analysis. It can minimize the operational error and maximize the margin hyperplane, as a result it will maximize the classification performance [12].

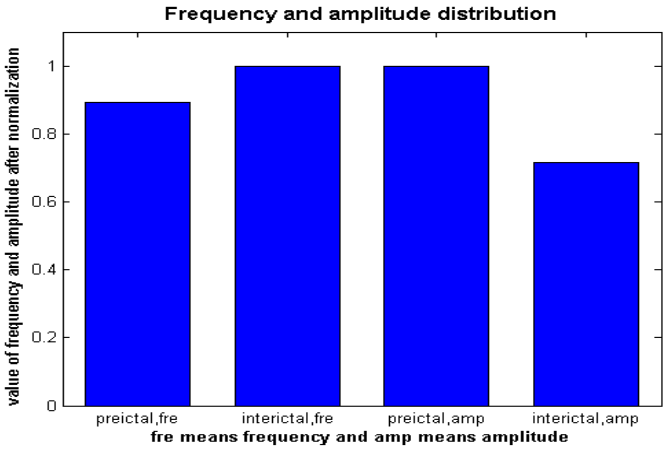


Fig. 4. Frequency and amplitude distribution on Frontal lobe EEG signals. Preictal signals carries lower and higher distribution for frequency and amplitude, respectively, whereas interictal signals has reversed order.

Let us consider binary classification problem with the training set $\{x_i\}_{i=1}^N$, along with corresponding targets $\{y_i\}_{i=1}^N$ where $x_i \in R^n$ and $y_i = \{1,-1\}, i = 1,...N$. Assume there exists a linear separable hyperplane $f(w,b) = \|w\|^2/2$ such that

$$w^T x_i + b \geq 1 \qquad \forall_i : y_i = 1$$

$$w^T x_i + b \leq -1 \qquad \forall_i : y_i = -1$$

where the decision function [11] is:

$$f(x) = sign\left[w^T g(x) + b\right] \tag{3}$$

where w is the weight vector and $g(x)$ is the mapping function of x and b is a bias.

The optimization problem of LS-SVM can be solved in the following way:

$$\min_{w,b,\xi} \frac{1}{2}\|w\|^2 + \frac{C}{2}\sum_{i=1}^{N}\xi_i^2 \tag{4}$$

subject to $y_i((w^T x_i) + b) = 1 - \xi_i, i = 1...N$.

where ξ_i are the slack variables, it represents the upper bound training error and $C > 0$ trades off margin size and training error.

Lagrange multipliers α_i can be introducing to solve optimization problem:

$$\min_{w,b,\xi,\alpha} \frac{1}{2}\|w\|^2 + \frac{C}{2}\sum_{i=1}^{N}\xi_i^2 - \sum_{i-1}^{N}\alpha_i\left\{1 - y_i\left[w^T g(x_i) + b\right] - \xi_i\right\} \tag{5}$$

subject to $\alpha_i \geq 0; i = 1...N$.

For non-linear problem the kernel function is introduced in the above problem. Therefore, the decision function can be defined:

$$f(x) = sign\left[\sum_{i-1}^{N} \alpha_i y_i K(x, x_i) + b\right] \quad (6)$$

where $K(x, x_i)$ is a kernel function.

Radial basis function (RBF) kernel is used in our experiments and this function can be defined as:

$$k(x, x_i) = \exp\left(-\|x - x_i\|^2 / 2\sigma^2\right) \quad (7)$$

where σ controls the width of the RBF function.

III. Experimental Results

We have classified the preictal and interictal EEG signals extracted from Frontal and Temporal lobe using LS-SVM classifier with RBF kernel where the value of regularization parameter and kernel parameter are 10 and 6 respectively. In our experiment, we have used 150 signals from preictal data and 882 signals from interictal data of three minutes EEG recording signals from dataset [7] and applied 4-fold cross validation for classification purpose. After feature extraction, we randomly sort the preictal and interictal data for the 1032 signals for each lobe either Frontal or Temporal. For unbiased classification results, we split the whole 1032 signals into four subsets; in each subset we randomly select 75% samples for the training and the remaining 25% for the testing. The sensitivity, specificity, and accuracy are defined as:

$$Sensitivity = (TP/(TP + FN)) \times 100 \quad (8)$$

$$Specificity = (TN/(TN + FP)) \times 100 \quad (9)$$

$$Accuracy = ((TP + TN)/(TP + TN + FP + FN)) \times 100 \quad (10)$$

where TP and TN represents the total number of detected true positive events and true negative events respectively. The FP and FN represent false positive and false negative respectively.

TABLE 1. SENSITIVITY, SPECIFICITY, AND ACCURACY FOR DIFFERENT FEATURES OF PREICTAL AND INTERICTAL EEG SIGNALS FROM FRONTAL AND TEMPORAL LOBE.

Kernel	*Lobe*		*Existing[8]*	*Proposed*
RBF	Frontal	SEN	47.88	97.05
		SPE	85.67	92.73
		ACC	79.94	93.12
	Temporal	SEN	55.20	94.00
		SPE	86.45	93.83
		ACC	86.14	93.79

It can be observed from Table 1 that the proposed method outperforms the state-of-the-art method [8] in terms of sensitivity, specificity, and accuracy. In terms of three different classification criteria, proposed method performs consistently and the classification performances are 93.12% and 93.79% from Frontal lobe and Temporal lobe signals respectively, whereas, the technique [8] does not perform consistently and the classification of performances are

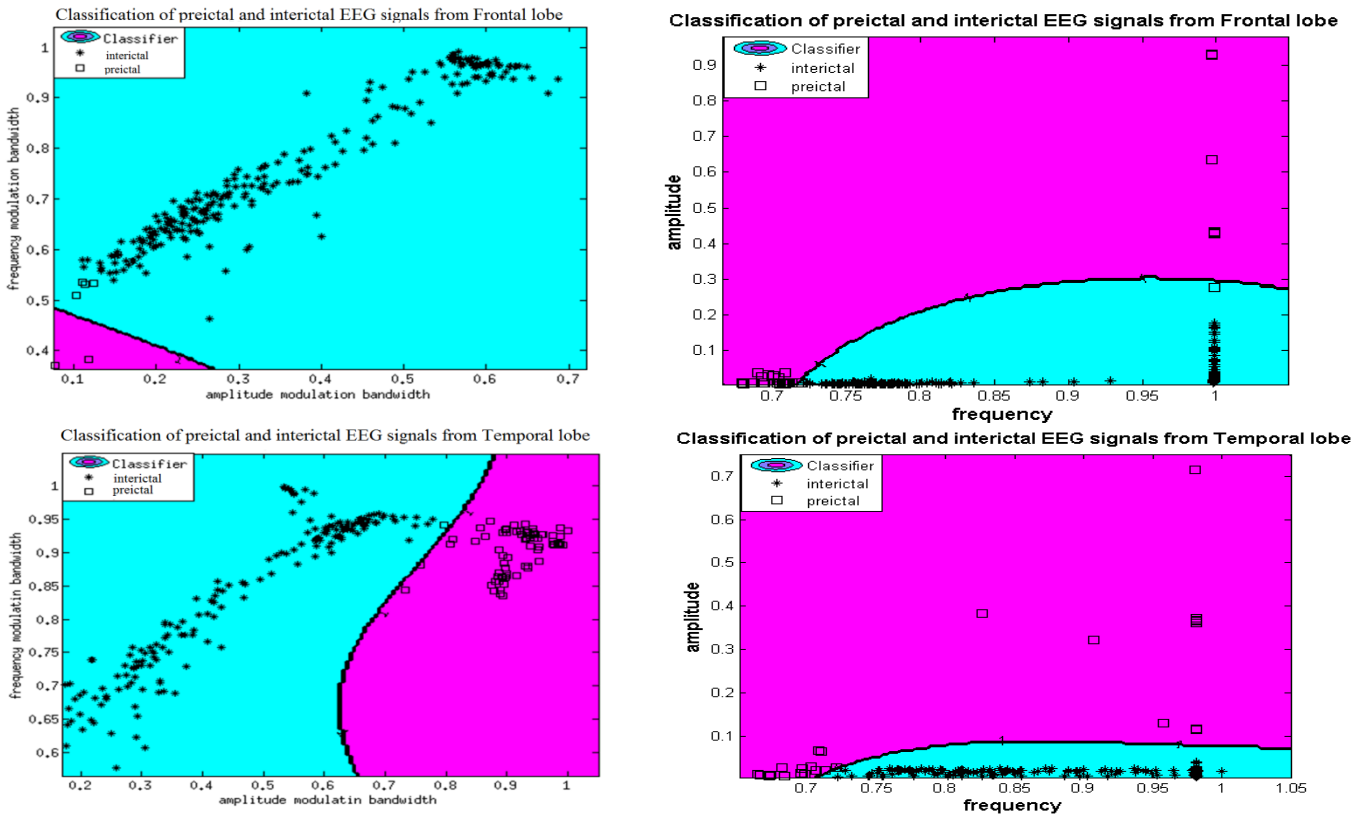


Fig. 5. Classification of preictal and interictal EEG signals for seizure prediction. First row represents the classification of preictal and interictal EEG signals from Frontal lobe for the first subset of testing using the state-of-the-art method against the proposed method while second row represents the classification of preictal and interictal EEG signals from Temporal lobe for the third subset of testing using the state-of-the-art method against the proposed method.

79.94% and 86.14% from Frontal lobe and Temporal lobe signals, respectively. The classification consistency among three different criteria is very important for proper diagnosis.

Fig. 5 shows the visual classification comparison using the proposed methods and the state-of-the-art method [8] for better understanding. We have generated two images in Fig. 5 from Frontal lobe for the first subset of testing set and obtained the classification accuracy 81% by the existing technique [8] (see top left of Fig. 5) and 94.18% by the proposed method (see top right of Fig. 5). On the other hand, state-of-the-art method provide 79.84% accuracy (see bottom left of Fig. 5) and proposed method also provide 95.34% accuracy (see bottom right of Fig. 5) for third subset from Temporal lobe signals. Moreover, visually we can see from Fig. 5 that the existing method has more miss-classified preictal signals compared to that of the proposed method. Thus, it can be concluded from Fig. 5 that the proposed methods outperform the state-of-the-art method using the dataset in [7].

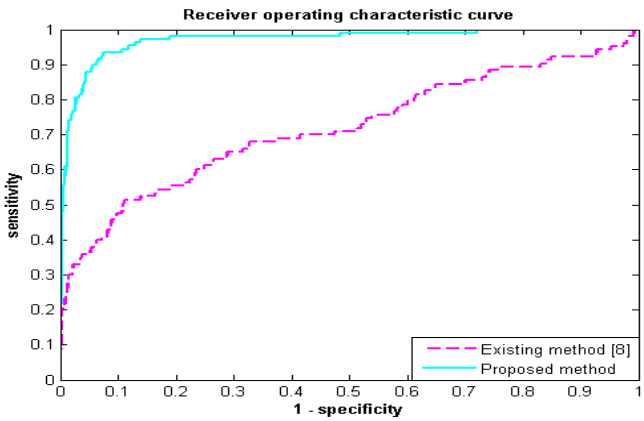


Fig. 6. The receiver operating characteristics (ROC) curves of the first subset of training and testing EEG signals by proposed method against the state-of-the-art method using LS-SVM with RBF kernel from Temporal lobe.

The performance of the LS-SVM is evaluated by *receiver operating characteristics* (ROC) plot is shown in Fig. 6. ROC illustrates the performance of a binary classifier system where it is created by plotting the fraction of true positives from the positives i.e., true positive rate (TPR) vs. the fraction of false positives from negatives i.e., false positive rate (FPR) with various threshold settings. TPR is known as sensitivity, and FPR is one minus the specificity or true negative rate. Fig. 6 demonstrates that the proposed method is carrying good classification results than that of the state-of-the-art method [8] using dataset [7] from Temporal lobe of the first subset of training and testing EEG signals.

IV. CONCLUSION

The progress of a reliable seizure prediction and detection could improve the quality of life of an epileptic patient significantly. Hence, it is important to detect forth-coming seizure to stop the ictal event and its associated consequences. For this purpose, we need to properly classify preictal and interictal EEG signals with good sensitivity and specificity. In this paper, we present a method using DWT to decompose the EEG signals into different bands and then extract features from the gamma band by exploiting temporal correlation for detection of preictal EEG signal. The results show that the proposed method offers a good combination of sensitivity and specificity, thus making the classification of preictal and interictal EEG signals feasible. Correct detection of preictal signals leads to seizure prediction by removing noisy classification results through regularization with decision function which is the purpose of our future work.

REFERENCES

[1] R. S. Fisher, W. van Emde Boas, W. Blume, C. Elger, P. Genton. P. Lee, and J. Jr. Engel," Epileptic seizures and epilepsy: definitions proposed by the International League Against Epilepsy (ILAE) and the International Bureau for Epilepsy (IBE)," Epilepsia, 46(4), 470-472, 2005.

[2] P. Kwan, and M.J. Brodie," Refractory epilepsy: mechanisms and solutions," Expert Review of Neurotherapeutics, 6(3):397-406, 2006.

[3] J. Rasekhi, M. R. K. Mollaei, M. Bandarabadi, C. A. Teixeira, and A. Dourado," Preprocessing effects of 22 linear univariate features on the performance of seizure prediction methods," Journal of Neuroscience Methods, vol. 217, 9–16, 2013.

[4] J. R. Williamsona, D. W. Blissa, D. W. Brownea, and J. T. Narayananb," Seizure prediction using EEG spatiotemporal correlation structure," Journal of Epilepsy & Behavior, vol. 25(2), 230–238, 2012.

[5] F. Duman, N. Ozdemir, and E. Yildirim," Patient Specific Seizure Prediction Algorithm Using HilbertHuang Transform," IEEE-EMBS International Conference on Biomedical and Health Informatics, Hong Kong and Shenzhen, China, 2012.

[6] L. Chisci, A. Mavino, G. Perferi, M. Sciandrone, C. Anile, G. Colicchio, and F. Fuggetta," Real-Time Epileptic Seizure Prediction Using AR Models and Support Vector Machines," IEEE Transactions on Biomedical Engineering, 57(5), 2010.

[7] EEG Data set from Epilepsy Center of the University Hospital of Freiburg, http://epilepsy.uni-freiburg.de/freiburg-seizure-prediction-project/eeg-database, visited date: June 10, 2012.

[8] V. Bajaj and R. B. Pachori, "Classification of Seizure and Non-seizure EEG Signals using Empirical Mode Decomposition," *IEEE Transactions on Information Technology in Biomedicine*, vol. 16(6), Nov 2012.

[9] "Epilepsy data: a few small files (text format)," http://epileptologie-bonn.de/cms/front_content.php?idcat=193&lang=3&changelang=3, visited date: April 28, 2012.

[10] Y. Song, and J. Zhang," Automatic recognition of epileptic EEG patters via Extreme Learning Machine and multiresolution feature extraction," Elsevier Journal of Expert Systems with Application, vol. 40, 5477-5489, 2013.

[11] J. A. K. Suykens and J. Vandewalle," Least Squares Support Vector Machine Classifiers," Neural Processing Letters, vol. 9(3), 293–300, 1999.

[12] K. D. Brabanter, P. Karsmakers, F. Ojeda, C. Alzate, and J. D. Brabanter, K. Pelckmans, B. De Moor, J. Vandewalle, J.A.K. Suykens, LS-SVMlab Toolbox User's Guide, version 1.8, Katholieke Universiteit Leuven, August 2011.

[13] H. Adeli, Z. Zhou, and N. Dadmehr,"Analysis of EEG records in an epileptic patient using wavelet transform," Journal of Neuroscience Methods, vol. 123, 69–87, 2003.

[14] S. Sanei and J.A. Chambers, "EEG Signal Processing," Centre of Digital Signal Processing, Cardiff University, UK, John Wiley & Sons, Ltd., 2007.

[15] X. Zhang, W. Diao and Z. Cheng, " Wavelet Transform and Singular Value Decomposition of EEG Signal for Pattern Recognition of Complicated Hand Activities," *Digital Human Modeling, Lecture Notes in Computer Science*, vol. 4561, 294-303, 2007.

An Enhanced Compression-Based Energy Efficient Data Communication Scheme for Wireless Sensor Networks

S. A. Ahsan Rajon *and* Md. Mahbubur Rahman
Computer Science and Engineering Discipline, School of Science, Engineering and Technology,
Khulna University, Khulna-9000, Bangladesh
ahsan.rajon@gmail.com, mahbubcse@yahoo.com

Abstract— **In this paper, we propose an effective data communication scheme for wireless sensor networks. The data communication approach makes efficient use of data compression to achieve energy efficiency. The main contribution of the proposed approach is, it makes use of the simple but efficient representation of value pairs to attain the compression. It is also novelty of our approach that, proposed scheme employs cluster coordinators for balancing energy consumption by assigning the inter cluster communications to the cluster heads and the intra cluster communications to cluster coordinators. Since, providing flexible service using wireless sensor network is challenging because they usually contain a large number of resource constraint sensing nodes with small memory, lower processing capability and extremely limited battery life, the scheme is considered to be simple, computationally feasible and overall resource requirement is minimum. It is also a mentionable feature of our research that, besides consideration of resource constraints of individual sensor nodes into concern, the proposed compression based compression scheme takes the overall routing scheme into consideration which ensures efficient and minimum energy consumption of WSNs.**

keywords: ***Wireless Sensor Networks, Data Compression, Resource Constraints, Power Consumptions, Data Communication, Routing, Cluster Coordinators.***

I. Introduction

Wireless Sensor Networks (WSN) have attracted lot of attention due to their multidimensional applications in both civilian and military arenas. Wireless Sensor Networks (WSN) usually consists of a large number of ultra small low-cost battery-powered devices that have limited energy resources, small computational capability, limited memory, and shorter communication capacities [1], [13]. According to different applications such as battlefield reconnaissance and homeland security monitoring, agricultural field monitoring, fire detection WSNs are often deployed in a vast terrain to detect events of interest and deliver data reports over multi-hop wireless paths to the sink. Data engineering and management is essential for these mission-critical applications to work in unattended and even hostile environments. WSNs can be used for data collection purposes in situations such that environmental monitoring, habitat monitoring, surveillance, structural monitoring, equipment diagnostics, disaster management, and emergency response [1], [2], [5], [13], [28].

The main contributions of this paper are-

Firstly, it proposes a computationally simple and energy efficient data compression based communication scheme where the workload distribution is more or less balanced and consequently, in spite of using the energy of any specific sensor especially the event reporting sensors at a greater level, we propose to achieve the total communication at a hop by hop basis, which essentially ensures robust data delivery.

Secondly, this paper employs the novel concept of cluster coordinator which is aimed to minimize the communication overheads of cluster heads by administering the intra-cluster communications and managements while inter-cluster and cluster-skin communications are supposed to be managed by cluster heads. For building the knowledge-base cluster coordinators plays a vital role and thus facilitates the entire compression process.

Thirdly, this paper employs the value-pair comparison based scheme which facilitates simple but effective solution in compressing the data.

Moreover, since proposed data compression approach is computationally simple and integration of cluster coordinators balances the energy consumption, overall network lifetime is also increased.

This paper is organized into six sections. In the related literature section, we provide elementary concepts of sensor network and data routing. Recent researches regarding data compression in wireless sensor networks are presented in the section three. A detailed overview of our proposed scheme has been illustrated in section three. Section four is devoted to present an extensive mathematical model and empirical analysis of the proposed scheme. This paper concludes with section five specifying future directions regarding proposed scheme.

II. Related Research

Data compression in wireless sensor networks may be effective only at that time when the energy consumption for compressing the data along with sending the compressed data is significantly less than transmission of non-compressed data. Moreover, since generic data compression algorithms are aimed at saving storage, not energy, most of the traditional data compression algorithms fail to satisfy the requirements for

wireless sensor networks. Consequently, it requires development of sophisticated energy consumption aware data compression algorithm for wireless sensor networks in order to attain power savings. Applying specific energy efficient routing mechanism along with effective data compression for WSNs may offer best role in maximizing the energy savings in Wireless network and thus enhance the overall lifetime of the network.

A. Compression of Short Text on Embedded Systems

The recent literature regarding short text compression titled "Compression of Short Text on Embedded Systems" by Rein et al. [25] proposes a low-complexity version of PPM (Prediction by Partial Matching). A hash table technique with one-at-a-time hash function is employed in this method to design the data structure for data and context model. They use statistical models with hash table lengths of 16384, 32768, 65536 and 131072 elements requiring two bytes for each element, which result an allocation of 32, 64, 128 and 256 Kbytes of RAM respectively. If this memory requirement may be substantially decreased, we may achieve more efficient compression and hence may make the scheme usable to even very low-quality cellular phones. [24].

B. Pipelined In-Network Data Compression (PINCO)

Pipelined In-Network Data Compression [20] is an elementary approach sophisticated for wireless sensor networks. The main underlying idea of PINCO is sharing common parts of the data. In the aggregation nodes, some technique is applied in combining the common parts of the correlated data which achieves the compression. The more consecutive common sensing data have the more bits can be aggregated using the common bits (known as prefix) and thus the better the compression. The format of sensed data is considered as <sensing data, node ID, time-stamp> before the compression while, after compression it is considered as <shared prefix, suffix list, node ID list, time-stamp list>. The revealed common part of the data is contained into the shared prefix part and the remaining parts are presented as a list of suffix. Corresponding node ids and timestamps are also presented as corresponding list. When the sensing data maintains high correlation, the approach is efficient whereas the compression ratio becomes poor for data having rare common parts. [19], [20]

C. Simple Huffman Algorithm

The main idea of simple Huffman algorithm is to map each symbol with some binary representation ensuring use of shorter bit-length codes for frequently appearing symbols than those occur rarely. For any sensor node, the measured data m_i is converted into its binary representation r_i comprising of R bits by an analog-to-digital converter (ADC), where R is the resolution of ADC. The compression algorithm computes $d_i=r_i—r_{i—1}$ for each new acquisition m_i. For computing d_0, it is assumed that, r_{-1} is equal to the central value among 2^R possible discrete values. Compression is performed by the entropy coder by encoding the differences d_i more compactly based on their statistical characteristics. A bit sequence bs_i composed of two parts $s_i|a_i$ is used to represent d_i, where s_i codifies the number n_i of bits needed to represent d_i and a_i is the representation of d_i. n_i is computed as $n_i = \lceil log_2(|d_i|) \rceil$ for any non-zero value of d_i ; while, $n_i=0$, for any $d_i=0$; Code s_i corresponds to a variable length symbol generated from n_i using Huffman coding. In [22], Marcelloni and Vecchio have also presented a Huffman variable length codes used for their research. The a_i part of the bit sequence bs_i is a variable-length integer code generated as follows: [22]

1) if $d_i>0$, a_i corresponds to the n_i low-order bits of the two's complement representation of d_i;

2) if $d_i<0$, a_i corresponds to the n_i low-order bits of the two's complement representation of (d_{i-1});

3) if $d_i=0$, s_i is coded as 00 and a_i is not represented.

The main disadvantage of Huffman Encoding is that, for optimal encoding, generation of histogram is required for each presented dataset. As a result, for each occasion, a new dictionary is generated. In order to decode the compressed data, the same dictionary used when compressing is required again. Thus, it would be necessary to transmit the dictionary with the encoded dataset, which obviously increases the transmission overhead [27].

D. LZW for Sensor Nodes (S-LZW)

LZW for Sensor Nodes (S-LZW) proposed by Sadler and Martonosi is based on the famous Lempel-Ziv-Welch (LZW) compression algorithm. This lossless compression scheme is actually a dictionary based compression approach. Original Input bit stream is divided into fixed sized blocks (*e.g.* 512 bytes) and then each block is compressed individually. For each new block, the dictionary is re-initialized with the codes (256 entries) representing standard character set. With each occurrence of new string in the bit stream, the dictionary is updated. Since, this frequent update may result into memory overload, proposed measure includes getting the dictionary frozen and using it as-is for encoding remainder of the data or resetting it and starting from scratch. Apart from LZW, a mini-cache has been introduced in S-LZW which is a hash-indexed dictionary of size N, where N is a power of 2, that stores recently used and created dictionary entries with an aim to take the underlying advantage of the repetitiveness characteristics of sensor data [21].

E. Simple Data Compression Algorithm

In [26], authors have proposed a re-compression technique for the base station node for saving of storage. Rather than handling with individual data item as tuple of node identifier and data itself, base station considers a compressed group of items composed with a common node-group identifier and followed by measurement data of each sensor node. This approach is effective for only storage optimization for the base stations while it has hardly any benefit in energy optimization for the individual sensor nodes and thus increasing life time and so forth.

III. Enhanced Compression Based Data Communication Scheme For WSNs

The main idea of proposed data communication scheme is to utilize the effectiveness of data compression along with the selection of appropriate routing scheme which facilitates the

maximum advantage of compression. The proposed compression scheme is mostly based on three interrelated steps.

The proposed data compression scheme is based on the preprocessing of two fold clustering scheme which employs cluster coordinators along with cluster heads for achieving the compression.

At the network initialization step, it is assumed that, a broadcasting scheme is adopted by each node which informs each node about its residual energy. Neighbors are also identified by each node. Neighbors are defined as those nodes which is reachable at the radio range of each node. This neighbor detection is particularly important for cluster formation. Whenever the detection is done, it builds an information table with its residual energy, neighbor nodes and corresponding residual energy. There exists other parameter regarding number of rounds it is acting as cluster head (initially zero), number of rounds it is acting as cluster coordinator (initially zero), current cluster head id, current cluster coordinator id. There are threshold values including Hard and Soft Threshold (as in TEEN).

Cluster coordinator is aimed to minimize the energy dissipation of nodes of each cluster. Cluster coordinator is selected as the node which has the maximum residual energy in that cluster other than cluster head. The main motivation of incorporating cluster coordinator is to balance the load distribution between inter cluster communication and intra cluster communication. Based on the distance threshold and percentage of coordinator threshold, there may be more than one cluster coordinator.

Cluster Heads perform the inter-cluster communication whereas cluster coordinators perform the intra cluster tasks. Since, inter cluster communication and cluster head to base station communication requires greater power than that of inner cluster communication, we are interested in proposing two level of cluster leaders namely cluster coordinators and cluster heads. The selection mechanism of cluster head is the same as proposed in LEACH. For cluster head selection, it also imposed that, any node currently engaged as cluster coordinator may not become cluster head for that round because this would lead to redundant exchange of the entire cluster information to the next cluster assistant. However, interchange between the cluster head and the cluster coordinator is supported. Cluster coordinators also perform as sensing node at the same time. Thus, the task of cluster heads is simply relaying the information received from the cluster coordinators to other cluster heads or to the sink or base station. All types of queries are also communicated through the cluster heads, however the actual processing and data gathering is done by the cluster coordinators. Cluster heads may gather and decide on data received only from another cluster head or cluster coordinator.

In first step of data communication, each sensing node participates into the data compression process. When the nodes are deployed, initial data are communicated without any compression. With the course of time, the nodes develop a dictionary with the sensed data dynamically until a threshold reaches. This dictionary building process synchronizes the same dictionary entry into the cluster coordinators. Initially each node is considered to have a blank dictionary. The dictionary is supposed to map a combination of two consecutive values (sensor readings) onto a single code. Whenever any node has any sensed data (α_j), it compares it with its just previous reading (α_i). If this combination of data (α_i and α_j) doesn't exist on the dictionary, the value couple is inserted into the dictionary as a new entry. At the same time a data packet is also sent to its cluster coordinator informing the newly explored dictionary entry and its codeword. On each cluster coordinate end, whenever it receives any request for updating its dictionary, it updates the same and also informs it to its sensing nodes. That is, dictionary of the cluster coordinators and individual nodes is synchronized at the same time. The same process is applied for cluster heads, where each cluster head builds a new dictionary for reception of data from other cluster head and also synchronizes the same with other cluster heads. Since, the cluster heads handle only the communication between the base station and cluster head and communication among the cluster heads, it reduces the overhead of communication among the individual nodes to cluster heads. Overhead of the communication among the sensor nodes *i.e.* intra-cluster communication is maintained by the cluster coordinators. Since, for cluster heads, communicating with the base station and other cluster heads requires data transmission towards longer distance than communication between the sensing nodes within the cluster, the probability of losing energy for cluster heads is much higher than that individual sensing nodes. When a single cluster head is to maintain both type of data communication, it obviously reduces the overall lifetime of the network. But, disseminating the intra cluster and inter cluster data communication to cluster coordinators and cluster heads essentially prolongs the network lifetime extending the uptime of cluster heads. It is worth to note that, we use independent dictionary on cluster heads and sensing nodes (with cluster coordinator) because, maintaining a single dictionary which is expected to be built on the fly adds the extra overhead of synchronizing that with all the elements of the network that results energy consumption into a greater level causing early death of the network. Separating inter cluster and intra cluster communication and providing individual dictionary for each cluster thus minimizes energy consumption a lot.

On the basis of the network architecture, maximum number of entries into the dictionary is agreed upon. For our experiments, we consider maximum 256 entries (*i.e.* 256 pairs of values) into the dictionary. Whenever the threshold limit exceeds, a packet informing the nodes regarding the completion of the dictionary building phase is sent to the nodes.

Whenever any node sends data to the cluster head, it maintains a certain packet format. The first entry comprise of *node_id* followed by the number of sensed data. Then for each data or value pair, *dictionary_flag* and *dictionary_entry* are placed. A *dictionary_flag* 0 indicates that, the value following the *dictionary_flag* is a single data; while a 1 as the *dictionary_flag* indicates that, *dictionary_entry* is the dictionary code for a value pair.

Whenever the cluster head have the dictionary formed, it is communicated with the sink. Since, the sink is supposed to have enough power and in most cases, it possesses enough computational capability to process huge data, maintaining separate dictionary for each cluster head is computationally feasible. Moreover, since, maintaining a single dictionary for all the cluster head requires communicating the dictionary to all the cluster head and ultimately to each node, the overall energy consumption may not be reasonable and may even result into network failure as occurs in adaptive Huffman coding. The second part includes the cluster heads where the received data from the nodes are processed to be transmitted to the sink. Whenever the cluster heads communicate the data with the sink, we adopt the pipelined in network compression approach [20]. It is notable that, the data part includes an amalgamation of both compressed and un-compressed data. Since, pipelined in network compression approach considers binary representation, it works very frequently with the data received from the first phase of the proposed scheme. .

IV. Analysis of Proposed Approach

In this section we provide a mathematical model for analysis of the energy consumption and coding overhead of our proposed scheme.

Let the total number of nodes be n.

Number of cluster = c

Number of Cluster Heads *i.e.* number of cluster coordinator is also c

Average number of nodes into each cluster = m

Distance from a sensing node to cluster coordinator = $d_{co\text{-}ss}$

Distance from a cluster coordinator node to cluster head= $d_{co\text{-}ch}$

Distance from two neighbor cluster heads = $d_{ch\text{-}ch}$; where $d_{ch\text{-}ch} > d_{co\text{-}ch}$ and $d_{ch\text{-}ch} > d_{co\text{-}ss}$

Average energy required to transmit 1 bit data among the sensing nodes = E_{ss}

Average energy required to transmit 1 bit data between the cluster head and cluster coordinator = $E_{ch\text{-}co}$

Average energy required to transmit 1 bit data among the cluster heads = $E_{ch\text{-}ch}$

A. Energy consumption for construction of the dictionary

Let the threshold for dictionary entry = d

Also assume that, the d entries are contributed by all the c clusters collectively and also assume that each cluster participates forming p entries into the dictionary. This p entries may be contributed by q sensors (considering $p>=q$).

Thus total energy consumption per cluster (E_{ce}) = Energy consumption for transmitting p value-pairs to the cluster coordinator (E_{co-ss_i}) + Energy consumption by cluster coordinator for reception of p value-pairs from the sensing nodes (E'_{ch-ch_i}) + Energy consumption for broadcasting p value-pairs (E_{co-ss_i}) by cluster coordinator (to be received by m sensor nodes plus the cluster head).

That is, Total energy consumption per cluster E_{ce} for intra cluster communication is

$$E_{ce} = \sum_{i=1}^{m}(d_{co-ss_i} \times q_i \times E_{co-ss_i}) + \sum_{i=1}^{m}(d_{co-ss_i} \times q_i \times E'_{co-ss_i}) + p\sum_{i=1}^{m+1} d_{co-ss_i} \times E_{co-ss_i}$$

Cluster head receives the newly formed dictionary entries from its cluster coordinator.

Other clusters also form the dictionary among them. This is synchronized among the cluster heads in the same way as that of the sensor nodes. Consider that each cluster head contributes r entries into the dictionary composed of s elements in total.

Thus total energy consumption per cluster head for inter cluster communication (E_{ch}) = Energy consumption for transmitting r value-pairs to other cluster heads (E_{ch-ch_i}) + Energy consumption by cluster heads for reception of r value-pairs from the sensing nodes (E'_{ch-ch_i}) + Energy consumption for broadcasting p value-pairs (E_{ch-ch_i}) by cluster coordinator (to be received by m sensor nodes plus the cluster head).

Energy consumption per cluster head for inter cluster communication E_{ch} is

$$E_{ch} = \sum_{i=1}^{c}\left(d_{ch-ch_i} \times r_i \times E_{ch-ch_i}\right) + \sum_{i=1}^{c}\left(d_{ch-ch_i} \times r_i \times E'_{ch-ch_i}\right) + s \times \sum_{i=1}^{c+1} d_{ch-ch_i} \times E_{ch-ch_i}$$

Now, while any node has sensed data, it first compares the data with its knowledgebase. Then for each match of value pair, the codeword is send. While sending codeword, another flag added indicating whether that contains a value-pair or codeword.

If there are a total of l number of readings, sensing nodes would send data for $l/2$ times. Among them if l_1 and l_2 indicates value-pair and codeword respectively,

$$B = \sum_{i=1}^{l1}(b_{c_i} + b'_{c_i}) + \sum_{i=1}^{l2}(b_{2s_i} + b'_{2s_i})$$

For uncompressed data, there would have l number of data transmission.

$$B' = \sum_{i=1}^{l}(b_{s_i} + b'_{s_i})$$

Here b_c is the bit length of codeword and b'_c includes the node identity while b_{2s_i} is the un-coded value-couple and b'_{2s} contains the value identity.

For uncompressed data communication approach, b_s and b'_s holds single data and identity of the node respectively. Here $l = l_1 + l_2$

For highly correlated data, $l_1 \to l/2$ and for low correlated data $l_2 \to l/2$

Moreover, for best case analysis, $b_{c_i} = (b_{s_i}/2)$ and for worst case, $b_{2s_i} = (b_{s_i}/2)$

$$\sum_{i=1}^{l1}(b_{c_i}) = \sum_{i=1}^{l}(b_{s_i}/2)$$

For codeword replaced tuples, besides of the node identifier another single bit is added indicating that, the tuple contains codeword. Moreover, for individual tuples node identifier is provided with each sensed data, while for proposed scheme, single node identifier is contained for value pairs. That is,

$$\sum_{i=1}^{l1} b'_{c_i} \leq \sum_{i=1}^{l} b'_{s_i}$$

And,

$$\sum_{i=1}^{l1} b'_{2s_i} \leq \sum_{i=1}^{l} b'_{s_i}$$

Consequently, for notions involving B and B', we may easily infer from the above expressions that,

$$\sum_{i=1}^{l1}(b_{c_i} + b'_{c_i}) + \sum_{i=1}^{l2}(b_{2s_i} + b'_{2s_i}) \leq \sum_{i=1}^{l}(b_{s_i} + b'_{s_i})$$

That is, $B \leq B'$

That is, the total bits required to transmit data in proposed scheme is less than that of uncompressed data.

Considering the energy consumption to be same for data reception and transmission for each sensing node, we find that

$$E_{ce} = 2\sum_{i=1}^{m}(d_{co-ss_i} \times q_i \times E_{co-ss_i}) + p \times \sum_{i=1}^{m+1} d_{co-ss_i} \times E_{co-ss_i}$$

Now, for compression of k readings on each sensing nodes, energy consumption is composed of energy consumption for each sensing nodes for sensing two consecutive readings (E_{ss_i}), comparison of the $k/2$ value-couple with the knowledgebase and encoding of data value pairs (E'_{comp_i}), and then sending the $k/2$ compressed data (indicating value-pairs) to the cluster coordinator (E_{co-ss_i}) while for uncompressed data communocation, energy consumption is composed of sensing individual k readings (E_{ss_i}), and then sending k individual data to the cluster head (E_{co-ss_i}).

For proposed scheme

$$E_{ce'} = \sum_{i=1}^{k}(E_{ss_i}) + \sum_{i=1}^{k/2} E'_{comp_i} + p'\sum_{i=1}^{k/2} d_{co-ss_i} \times E_{co-ss_i}$$

For uncompressed data communication approach,

$$E_{ce''} = \sum_{i=1}^{k}(E_{ss_i}) + p'\sum_{i=1}^{k} d_{co-ss_i} \times E_{co-ss_i}$$

Since, the energy required to transmit a single bit is much higher than that of the computational cost, from $E_{ce'}$ and $E_{ce''}$ we may conclude that,

$$E_{ce'} \leq E_{ce''}$$

That is, energy consumption for our proposed scheme is definitely lower.

Performance of the proposed scheme along with relevant schemes has been measured in terms of compression ratio.

$$Compression\ ratio = 1 - \left(\frac{bitsize\ after\ compression}{original\ size}\right)$$

Here size is usually measured in terms of bits.

The performance of our proposed scheme has been evaluated against the PINCO and S-LZW scheme. We have also evaluated the scheme considering both 256 entries and 512 entries as the threshold of the dictionary for each cluster head. We have also considered highly correlated data and low correlated data in out experiments. In figure 1, we have presented the compression ratio of the schemes for highly correlated data. Like other dictionary based schemes which usually perform slightly better for greater number of dictionary entries, our proposed scheme and S-LZW has been found better forming for 512 dictionary entries than 256 dictionary entries. Since, PINCO is not subjected to any dictionary mapping, varying the dictionary doesn't affect overall data compression. Moreover, our scheme and PINCO both exploits the correlation feature of data, performance of both the schemes for highly correlated data is better than low correlated data.

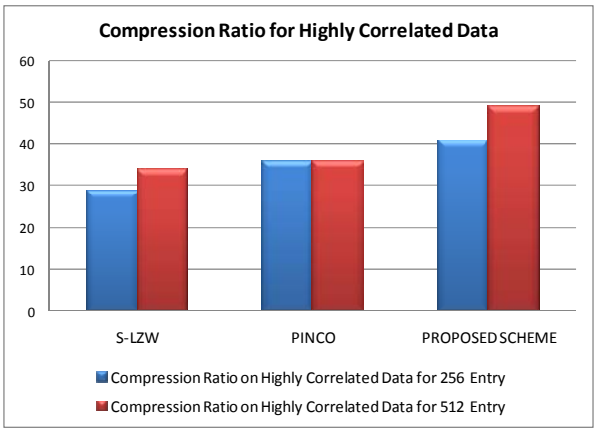


Fig. 1: Compression Ratio for highly correlated data

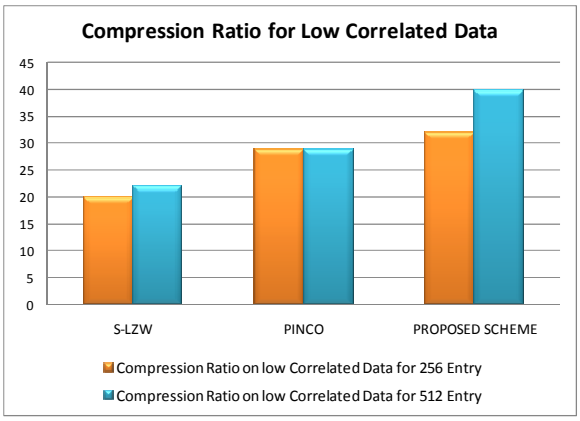


Fig. 2: Compression Ratio for low correlated data

V. Conclusion

In this paper, we have presented an overview of the existing researches on Sensor Network data compression approaches. This paper also proposes a compression technique which is also computationally simple and effective. We employ value-couple encoding in order to reduce the overhead of representing individual data. Our proposed scheme ensures balanced energy consumption and thus enhances the network lifetime. Future works may be dedicated to derive a more suitable way to minimizing the data transmissions involved in communicating the initial dictionary among the sensor nodes. Also the parameter of dictionary entry threshold may be standardized.

References

[1] Abdul Gani Khan, "A Comparative Analysis: Routing Protocols For Wireless Sensor Networks", *International Journal of Research in IT & Management (IJRIM)*, Volume 2, Issue 2 (February 2012)

[2] I.F. Akyildiz, W. Su, Y. Sankarasubramaniam, E. Cayirci, "Wireless sensor networks: a survey", *Computer Networks 38 (2002) 393–422, ELSEVIER Science*, 2002.

[3] T. Venu Madhav and N V S N Sarma "Energy Efficient Routing Protocol with Improved Clustering Strategies for Homogeneous Wireless Sensor Networks", *International Journal of Computer Applications* (0975 – 8887) Volume 38– No.8, January 2012

[4] A. Bari, A. Jaekel and S. Bandyopadhyay, "Maximizing the Lifetime of Two-Tiered Sensor Networks", In the Proceedings of *IEEE International Electro/Information Technology Conference,* 2006, 222–226.

[5] V.Chandrasekaran, Dr.A.Shanmugam, "A Review On Hierarchical Cluster Based Routing In Wireless Sensor Networks", *Journal of Global Research in Computer Science (JGRCS)*, Volume 3, No. 2, February 2012

[6] Ali Norouzi, Abdul Halim Zaim, "An Integrative Comparison of Energy Efficient Routing Protocols in Wireless Sensor Network", in *Scientific Research, Wireless Sensor Network,* 2012, 4, 65-75

[7] Shabnam Zabihi, Saeed Hadidi, Maedeh Khani, "A New Energy-Efficient Reliable Data Transfer Protocol for Data Transfer in WSN", *Journal of Basic and Applied Scientific Research*, 2(2)1972-1980, 2012.

[8] Mohammad Hossein Anisi, Abdul Hanan Abdullah, Shukor Abd Razak and Md. Asri Ngadi, "An Overview of Data Routing Approaches for Wireless Sensor Networks", in *Sensors*, 2012, 12, 3964-3996;

[9] A. Manjeswar and D.P. Agrawal, "TEEN: A Protocol for Enhanced Efficiency in Wireless Sensor Networks", in the Proceedings of 15th *International Parallel and Distributed Processing Symposium*, 2002, pp. 2009 – 2015.

[10] D. Braginsky, D. Estrin, "Rumor Routing Algorithm for Sensor Networks", in the Proceedings of the *First Workshop on Sensor Networks and Applications (WSNA)*, October 2002, Atlanta, GA, pp. 141-55.

[11] N. Sadagopan et al., "The ACQUIRE mechanism for efficient querying in sensor networks," Proceedings of the First International *Workshop on Sensor Network Protocol and Applications*, Anchorage, Alaska, May 2003, pp. 149 – 155

[12] W. Heinzelman, J. Kulik, and H. Balakrishnan, "Adaptive protocols for information dissemination in wireless sensor networks," Proceedings of the 5th Annual *ACM/IEEE International Conference on Mobile Computing and Networking (MobiCom'99)*, Seattle, WA, August 1999, pp.1-15.

[13] Kui Ren, Wenjing Lou, and Yanchao Zhang, "LEDS: Providing Location-Aware End-to-End Data Security in Wireless Sensor Networks", *IEEE Transactions on Mobile Computing,* vol. 7, no. 5, May 2008.

[14] S. Lindsey, C. Raghavendra,"PEGASIS: Power- Efficient Gathering in Sensor Information Systems," *IEEE Aerospace Conference Proceedings*, 2002, Vol. 3. No. 9-16, pp. 1125 1130.

[15] Intanagonwiwat, C. Govindan R. and Estrin,D. "Directed Diffusion: A Scalable and Robust Communication Paradigm for Sensor Networks". In Proceedings of the *Sixth Annual International Conference on Mobile Computing and Networks (MobiCOM 2000)*, August 2000, Boston, Massachusetts.

[16] Seokhoon Yoon, Abul K. Azad, Hoon Oh and Sunghwan Kim, "AURP: An AUV-Aided Underwater Routing Protocol for Underwater Acoustic Sensor Networks", in *Sensors*, 2012 , 12, 1827-1845.

[17] Nidhi Gautam, Prof Sanjeev Sofat, Prof Renu Vig, "Data Dissemination in Wireless Sensor Networks: A State-of-the Art Survey", *International Journal of Computer Networks (IJCN)*, Volume (4) : Issue (1) : 2012

[18] Mariam Faruque Sharif, Shalim Ahmed, Monjur-E-Mawla, Md.Roknuzzaman, and Kazi Chandrima Rahman, "An Efficient Data Aggregation Protocol Based on Dynamic Routing for Wireless Sensor Network", *International Journal of Computer and Information Technology,* IJCIT, ISSN 2078-5828 (PRINT), ISSN 2218-5224, Volume 02, Issue 02, 2012

[19] Hsing-Shao Liu, Chi-Cheng Chuang, Chih-Chung Lin, Ray-I Chang, Chia-Hui Wang and Ching-Chia Hsieh, "Data Compression for Energy Efficient Communication on Ubiquitous Sensor Networks", *Tamkang Journal of Science and Engineering*, Vol. 14, No. 3, pp. 245-254 (2011)

[20] Arici, T., Gedik, B., Altunbasak, Y. and Liu, L., "PINCO: A Pipelined In-Network Compression Scheme for Data Collection in Wireless Sensor Networks", Proceedings of 12th *International Conference on Computer Communications and Networks* (2003)

[21] C. M. Sadler and M. Martonosi, "Data compression algorithms for energy-constrained devices in delay tolerant networks," in Proc. of *SenSys'06: 4th International Conference on Embedded networked sensor systems,* 2006, pp. 265–278.

[22] Francesco Marcelloni and Massimo Vecchio, "A Simple Algorithm for Data Compression in Wireless Sensor Networks", *IEEE Communications Letters,* Vol. 12, No. 6, June 2008

[23] S. A. Ahsan Rajon and Md. Mahbubur Rahman, "Energy Efficient Routing in Wireless Sensor Network", *Annual Conference and Seminar on "Electronics and ICT for National Development*, *NSCE&T-2012*, October 3-4, 2012, Dhaka, Bangladesh.

[24] Md. Rafiqul Islam, S. A. Ahsan Rajon, "An Enhanced Scheme for Lossless Compression of Short Text for Resource Constrained Devices", *In the proceedings of 14th International Conference on Computer and Information Technology (ICCIT' 2011)*, December 22-24, 2011, Dhaka, Bangladesh.

[25] Stephan. Rein, F. Fitzek, M. P. G. Perucci, T. Schneider and C. Guhmann, "Low Complex and Power Efficient Text Compressor for Cellular and Sensor Networks", In 15th *IST Mobile and Wireless Communication Summit*, June 2006.

[26] Phayong Sornsiriaphilux, Dusit Thanapatay, Kamol Kaemarungsi and Kiyomichi Araki, "On Applying Simple Data Compression to Wireless Sensor Networks", *ANSCSE14* Mae Fah Luang University, Chiang Rai, Thailand, March 23-26, 2010.

[27] Andrew van der Byl, Robert Neilson, Richardt H. Wilkinson, "An Evaluation of Compression Techniques for Wireless Sensor Networks", IEEE AFRICON 2009, 23 - 25 September 2009, Nairobi, Kenya.

[28] S. A. Ahsan Rajon and Md. Mahnbubur Rahman, "Energy Efficient Routing in Wireless Sensor Network", *International Journal of Computer and Information Technology,* IJCIT, ISSN 2078-5828 (Print), ISSN 2218-5224 (Online), Volume 03, Issue 01, 2012, pp. 12-19

Performance of Some Parametric Test Statistics for Testing the Difference of Means of Two Skewed Populations

Shipra Banik
School of Engineering and Computer Science
Independent University, Bangladesh
Dhaka, Bangladesh
shipranbanik@yahoo.com.au

B.M. Golam Kibria
Department of Mathematics and Statistics
Florida International University
University Park, Miami, USA
kibriag@fiu.edu

***Abstract*—Right skewed data is available in the field of health science, biological science, epidemiological science and others. This paper assessed some existing procedures for testing the difference of means of two right skewed populations. A Monte Carlo simulation study has been conducted to compare the performance of the test statistics in the sense of empirical size and power of the selected tests. Based on the simulation study results, some good test statistics are recommended for practitioners. A real life health related example has been considered to illustrate the application of the techniques.**

***Index Terms*— Polymorphism, Power, Primary hypertension patients, Right skewed, Simulation, Size of the test, Statistical Inference**

INTRODUCTION

Researchers, such as biological scientists, health professionals, epidemiologists and others (for examples, see, Gregoire and Schabenberger [1], Andersson [2], Zhou and Dinh [3], Baklizi and Kibria [4], Banik and Kibria [5] and many others) are often interested in making inference about the difference of two important location parameters for two parent populations. To draw conclusions about the unknown parameter, we may consider either the confidence interval method or the hypothesis testing method. Confidence interval on the difference of two means have been appeared extensively in the literature (e.g. [3-5], Banik and Kibria [6] and many others). However, literature for testing the difference of means for two skewed populations is limited. Various researchers proposed different test statistics for testing the difference of two population means. Most of them are used normal theory to construct the test statistics. However, in our real life, a lot of health related data are skewed (e.g. [3], Almonte and Kibria [7], Shi and Kibria [8], Banik and Kibria [9] and others). It is known that the ordinary-*t* statistic suffers from problems when random samples come from skewed populations. To overcome this problem, different techniques have been appeared in the literature. For a detailed review of the available techniques, see Kibria and Banik [10].

In this paper, following [5], we propose some test statistics based on the modified pooled sample variance which is a function of sample mean absolute deviations. Since a theoretical comparison is not possible, a Monte Carlo simulation study has been chosen to compare the performance of test statistics. We have evaluated the performances of test statistics based on size and the power properties under both normal and skewed distributions. The organization of the paper is as follows. Section 2 reviews test statistics for testing the difference of means of two populations. A simulation study along with results discussion is presented in section 3. Real life health data is analyzed in section 4. Finally, some useful concluding remarks are given in final Section.

SOME TEST STATISTICS

In this section, some selected test statistics for testing the difference of two means will be discussed. Suppose $X_1, X_2, \ldots, X_{n1}$ and $Y_1, Y_2, \ldots, Y_{n2}$ are two independent random samples drawn from two populations with means μ_1 and μ_2 respectively. Let $\bar{x}$ and $\bar{y}$ are sample means and s_1^2 and s_2^2 are sample variances respectively.

Consider the following hypotheses for testing two means for two populations:

Null hypothesis: $H_0: (\mu_1 - \mu_2) = 0$ (1)

Alternative hypothesis: $H_1: (\mu_1 - \mu_2) \neq 0$ (2)

where

$$\mu_1 - \mu_2 = c \quad (3)$$

and c is a positive constant. When c=0, we get size (type I error rate (α)) and for c>0, we get powers (1-β, β is the Type II error rates) of the test statistic. In the following, we briefly discussed considered and our proposed two tests to test (1) against (2).

PARAMETRIC TESTS

Ordinary-t test

To test (1) vs. (2), the test statistic is

$$Z = \frac{\bar{x}-\bar{y}}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}, \quad \text{for large samples } (n>30)$$

$$t = \frac{\bar{x}-\bar{y}}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}}, \quad \text{for small samples } (n\leq 30)$$

where $\bar{x}$ and $\bar{y}$ are sample means, s_1^2 and s_2^2 are sample variances and n_1 and n_2 are sample sizes of two populations. For critical values of Z and *t* statistics, see standard *Z* and *t* tables respectively.

978-1-4799-3497-3/13/$31.00 ©2013 IEEE

Welch-Satterthwaite (WS) test

Welch [11] and Satterthwaite [12] proposed two test statistics for normal and skewed situations. Their proposed test statistics are given as follows:

When distribution is normal
The test statistic for testing (1) vs. (2) is defined as

$$t_{WS1} = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

where degrees of freedom(df) is calculated as
$df = \frac{(w_1+w_2)^2}{\left(\frac{w_1^2}{n_1-1}\right)+\left(\frac{w_2^2}{n_2-1}\right)}$, $w_1 = \frac{s_1^2}{n_1}$ and $w_2 = \frac{s_2^2}{n_2}$.

When distribution is non-normal

Cressie and Whitford [13] and Reed and Strak [14] show that performance of the t_{WS1} statistic can be much lower if random samples come from the skewed distribution. For this case, taking logarithm makes the distribution is more normal. If the distribution is non-normal, apply the t_{WS1} statistic to the log-transformed data and adjust the statistic (denoted by t_{WS2}) to its original scale. For critical values of the t_{WS2} statistic, use the standard *t* table.

Zhou and Dinh test

Zhou and Dinh [3] customized t-statistic (described as above) to obtain better performance when observations come from the skewed distribution. Their proposed test statistics for testing (1) vs. (2) are defined as below:

$$ZD1 = T_1^{-1}(t) = (\tfrac{1}{3}\hat{A})^{-1}\left\{1 + 3\tfrac{1}{3}\hat{A}(t - N^{-1}\tfrac{1}{6}\hat{A})\right\}^{\frac{1}{3}} - (\tfrac{1}{3}\hat{A})^{-1}$$
$$ZD2 = T_2^{-1}(t) = \left(2\tfrac{1}{3}N^{-\frac{1}{2}}\hat{A}\right)^{-1}\log\left\{2\tfrac{1}{3}N^{-\frac{1}{2}}\hat{A}\left(t - N^{-1}\tfrac{1}{6}\hat{A}\right) + 1\right\}$$
$$ZD3 = T_3^{-1}(t) = \left\{1 + 3\left(t - N^{-1}\tfrac{1}{6}\hat{A}\right)\right\}^{\frac{1}{3}} - 1$$

where $\hat{A} = \frac{(\frac{N}{n_1})^2 s_1^3\hat{\gamma}_1 - (\frac{N}{n_2})^2 s_2^3\hat{\gamma}_2}{\left\{(\frac{N}{n_1})s_1^2 + (\frac{N}{n_2})s_2^2\right\}^{3/2}}$, $\gamma_i = \frac{n_i}{(n_i-1)(n_i-2)}\sum_{j=1}^{n_i}\left\{\frac{X_{ij}-\bar{X}_i}{S_i}\right\}^3$, i=1,2, $N=n_1+n_2$, $\varepsilon_\varepsilon=\varphi(\alpha)$ is the cumulative distribution function of N(0,1) and $\hat{\sigma} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$. For critical values of the ZD1, ZD2 and ZD3 test statistics use the standard *Z* table.

Johnson test

Johnson [15] modified t-statistic using the first few terms of the inverse Cornish-Fisher expansion and the test statistic is defined as

$$t_{JOHN} = \frac{(\bar{x} - \bar{y}) + \left(\frac{\tilde{\mu}_3}{6N\tilde{\sigma}^2}\right) + \left(\frac{\tilde{\mu}_3}{3\tilde{\sigma}^4}\right)(\bar{x} - \bar{y})^2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

where $\tilde{\mu}_3 = \left(\frac{\tilde{\mu}_{31}}{n_1}\right) - \left(\frac{\tilde{\mu}_{32}}{n_2}\right)$ and $\tilde{\mu}_3$ is an estimate of the third central moment of the ith populations. For critical values of the t_{JOHN} statistic use the standard *t* table.

MAD t test

Following Banik and Kibria [5], we propose the following two test statistics for testing (1) vs. (2):

MADt test 1
The proposed test statistic 1 is defined as

$$t_{MADt1} = \frac{(\bar{x} - \bar{y})}{\tilde{S}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

where $df=n_1+n_2-2$ and $\tilde{s} = \sqrt{\frac{(n_1-1)MAD1+(n_2-1)MAD2}{n_1+n_2-2}}$ is the modified pooled samples standard deviation, $MAD1 = Median|x_{1i} - median(x_{1i})|$ is the median of absolute deviations from the first samples median and $MAD2 = Median|x_{2i} - median(x_{2i})|$ is the median of absolute deviations from the second samples median. For critical values of the t_{MADt1} statistic, use the standard *t* table.

MADt test 2
The proposed test statistic 2 is defined as

$$t_{MADt2} = \frac{(\bar{x} - \bar{y})}{\tilde{S}'\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

where $df=n_1+n_2-2$ and $\tilde{s}' = \sqrt{\frac{(n_1-1)MAD1*+(n_2-1)MAD2*}{n_1+n_2-2}}$ is the modified pooled samples standard deviation, $MAD1* = Mean|x_{1i} - median(x_{1i})|$ is the mean of absolute deviations from the first samples median and $MAD2* = Mean|x_{2i} - median(x_{2i})$ is the mean of absolute deviations from the second sample median. For critical values of the t_{MADt2} statistic, use the standard *t* table.

SIMULATION STUDY

A Monte Carlo simulation study is conducted to compare and evaluate the performances of selected test statistics defined in the previous section. We evaluate performance of test statistics based on empirical size and powers, which are calculated as the fraction of rejections of (1) out of 5000 replications. Empirical size and powers of selected statistics for testing (1) vs. (2) are calculated by (3) with c = 0.0, 0.05, 0.15, 0.35, 0.50, 0.75. We get size of the test (α) when c =0, otherwise powers of the considered test statistics. The most common 5% level of errors (α=0.05) is used to calculate size and powers of the test statistics. Sample sizes combinations are used: (8,8), (10,10), (15,8), (15,15), (20,15), (20,20), (40,30), (80,100).

978-1-4799-3497-3/13/$31.00 ©2013 IEEE

TABLE I: ESTIMATED SIZE OF SELECTED TESTS AT THE 5% SIGNIFICANCE LEVEL

Distributions	(n_1,n_2)	t	t_{WS1}	t_{WS2}	ZD1	ZD2	ZD3	t_{JOHN}	t_{MADt1}	t_{MADt2}
N(3,2) and N(2,1)	8,8	0.0300	0.0277	0.0223	0.2080	0.0486	0.0329	0.0303	0.0280	0.0280
	10,10	0.0297	0.0266	0.0211	0.2151	0.0457	0.0317	0.0303	0.0271	0.0271
	15,8	0.0309	0.0309	0.0289	0.2209	0.0394	0.0309	0.0309	0.0300	0.0266
	15,15	0.0306	0.0374	0.0357	0.2386	0.0389	0.0377	0.0303	0.0380	0.0380
	20,15	0.0366	0.0354	0.0323	0.2220	0.0320	0.0357	0.0369	0.0349	0.0349
	20,20	0.0371	0.0360	0.0311	0.2597	0.0374	0.0331	0.0371	0.0354	0.0366
	40,30	0.0323	0.0320	0.0317	0.2620	0.0343	0.0383	0.0323	0.0320	0.0334
	80,100	0.0326	0.0323	0.0317	0.2660	0.0334	0.0369	0.0323	0.0323	0.0389
G(0.25,1) and G(0.75,1)	8,8	0.0637	0.0594	0.0477	0.1069	0.0586	0.0677	0.0629	0.0474	0.0474
	10,10	0.0620	0.0586	0.0503	0.1071	0.0586	0.0623	0.0637	0.0471	0.0471
	15,8	0.0894	0.0740	0.0857	0.1069	0.0737	0.0894	0.0906	0.0740	0.0600
	15,15	0.0509	0.0503	0.0457	0.1043	0.0591	0.0489	0.0523	0.0406	0.0406
	20,15	0.0700	0.0651	0.0651	0.1146	0.0723	0.0651	0.0711	0.0571	0.0543
	20,20	0.0731	0.0686	0.0709	0.1151	0.0834	0.0683	0.0740	0.0589	0.0526
	40,30	0.0620	0.0500	0.0500	0.1166	0.0777	0.0451	0.0526	0.0417	0.0423
	80,100	0.0623	0.0523	0.0511	0.1034	0.0591	0.0543	0.0517	0.0537	0.0534

Note: t- Ordinary t, WS1- Welch-Satterthwaite 1,WS2- Welch-Satterthwaite 2, ZD1 – Zhou and Dinh 1, ZD2 – Zhou and Dinh 2, ZD3 – Zhou and Dinh 3, John- Johnson, MADt1 - MADt Method 1, MADt2 - MADt Method 2

SIMULATION PARAMETERS

Data are generated from the following two most common parent population distributions:

Symmetric distribution: Two random samples are generated from N(3,2) and N(2,1) with skewness 0.

Skewed distribution: Two random samples are generated from G(0.25,1) and G(0.75,1) with skewness 4.00 and 2.3094 respectively. Simulation results are presented in Tables I to III and discussed in the following section.

RESULTS AND DISCUSSION

Table I reports estimated sizes of the selected test statistics for normal and gamma distributions. In this table, columns 3-11 present proportion of times the null hypothesis (1) was rejected at the 5% significance level. We observed that for the normal distribution and for all sample sizes, the selected test statistics have reasonable size properties. It means that estimated sizes are close to the nominal size 0.05 except ZD1, which is suffering from size distortions. We found the same situation for gamma distribution. However estimated sizes are smaller than the nominal size for normal distribution while than for gamma distribution. In general, all tests have reasonable size properties except ZD1 when data generated from normal and gamma distributions. If we compare these test statistics based on the nominal size we can say that our proposed test statistics are performing better than the rest.

The estimated powers of test statistics for normal and gamma distributions are presented in Table II and Table III respectively. The first column gives values of c, the difference between two true mean values of populations. The entries in columns 3-11 represent the proportion of times H_0 was rejected at α=0.05 given (1) was false. From the Table II, it is clear that all tests have no acceptable powers (powers≥0.7) except ZD1 and for large sample sizes (n_1,n_2) = (80,100). However, ZD1 is not advisable due to lack of attaining nominal size. For normal population, ZD1 has the higher power while ZD3 has the lowest. Rest of test statistics has similar and reasonable good powers. In Table III we have reported estimated powers when data are generated from gamma distributions. All tests statistics show a noticeable improvement under gamma distribution as compare to normal distribution except, ZD1. Overall, t_{WS1}, t_{WS2}, t_{JOHN}, t_{MADt1} and t_{MADt2} statistics are performing very well compared to the rest when data are from skewed populations.

ILLUSTRATIONS OF THE TEST STATISTICS

Recently, Rossi et. al [16] and [5] studied total cholesterol measurements (mg/dl) for 133 primary hypertensive (PH) patients and 41 normotensive (NT) patients. Now, we want to test difference of means of PH and NT patients. Using test statistics (section 2), our target is to test whether PH patients have, on average, higher total cholesterol levels than NT patients. Summary statistics of the total cholesterol measurements (mg/dl) for 133 PH patients and 41 NT patients are presented in the Table IV. We found that two samples are

978-1-4799-3497-3/13/$31.00 ©2013 IEEE

TABLE IV DESCRIPTIVE MEASURES

Chol. meas(mg/dl)	n	Mean	Median	SD	Sk
PH patients	133	214.71	214	29.65	0.25
NT patients	41	193.19	194	33.04	0.56

positively skewed with skewed coefficients (Sk) 0.25 and 0.56 respectively. Fig.1 represents histograms and Q-Q plots of two considered series. It clearly shows us both samples are skewed. Given that two samples are skewed (Table IV), our plan is to see whether we can conclude that PH patients have, on average, higher total cholesterol levels than NT patients. It is well known for a typical hypothesis testing, a p-value is necessary to find whether one should accept or reject (1). The test statistics for testing the above claim and corresponding p-values for mean differences of two populations are tabulated in the Table V. We observed that by all test statistics, the population means are different at the 5% level of significance. On the basis of the Table V results we may conclude that PH patients have on average higher total cholesterol levels than NT patients. However, at the 1% significance level, ZD1, ZD2 and ZD3 produce insignificant results. We observed that the smallest p-values are obtained from our proposed test statistics.

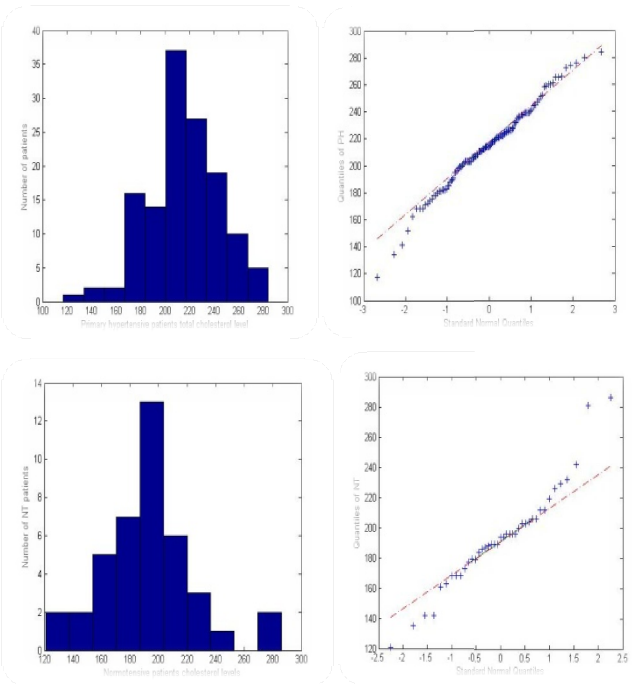


Fig.1 Histograms and Q-Q plots of total cholesterol measurements of PH and NT patients

TABLE V TEST STATISTICS AND P-VALUES

Test	Statistic values	p-value
t	3.7323*	0.0050
WS1	3.7323*	0.0050
WS2	3.7980*	0.0049
ZD1	2.2724*	0.0200
ZD2	2.2935*	0.0195
ZD3	2.0289*	0.0230
John	3.7288*	0.0050
MADt 1	29.6257*	0.0007
MADt 2	25.3108*	0.0006

Note: * indicates significant at the 5% level of errors

CONCLUSION

This paper considers some test statistics for testing differences between means of two skewed populations. We have evaluated performances of test statistics based on the size and powers by a simulation study under both symmetric and skewed conditions. Based on our simulation results, we recommended the *t*-test for the symmetric distribution and for the skewed distribution all tests to practitioners except *t*-test, ZD1 test and ZD2 test. A real life health related data has been considered to illustrate the applications of the techniques. We believe findings of this paper will be useful for the practitioners who want to select test procedures for testing differences of means for two skewed populations.

REFERENCE

[1] T.G. Gregoire and O. Schabenberger, "Sampling skewed biological populations: Behavior of confidence intervals for the population total", Ecology, 1999, pp.1056-1065.

[2] P.G. Andersson, "Alternative confidence intervals for the total of a skewed biological population", Ecology, 2004, pp.3166-3171.

[3] S. Zhou and P. Dinh, "Nonparametric confidence intervals for the one and two sample problems", Biostatistics, 2005, pp.187-200.

[4] A. Baklizi and B.M.G. Kibria, "One and two sample confidence intervals for estimating the mean of skewed populations: An empirical comparative study", Journal of Applied Statistics, 2009, pp. 601–609.

[5] S. Banik and B.M.G. Kibria, "A simulation study on some parametric confidence intervals for the difference of means of two skewed populations", IEEE Xplore, 2011, pp.65-70.

[6] S. Banik and B.M.G. Kibria, "On some parametric confidence intervals for the mean difference of two populations", Independent Business Review, 2012, pp.11-26.

[7] C. Almonte and B.M.G. Kibria, "On some classical, bootstrap and transformation confidence intervals for estimating the mean of an asymmetrical population" Model Assisted Statistics and Applications, 2009, pp.91-104.

[8] W. Shi and B.M.G. Kibria, "On some confidence intervals for estimating the mean of a skewed population", International

978-1-4799-3497-3/13/$31.00 ©2013 IEEE

TABLE II: ESTIMATED POWERS OF SELECTED TESTS FOR N(3,2) AND N(2,1) WITH SKEWNESS 0.0 AT THE 5% SIGNIFICANCE LEVEL

c	(n_1,n_2)	t	t_{WS1}	t_{WS2}	ZD1	ZD2	ZD3	t_{JOHN}	t_{MADt1}	t_{MADt2}
0.05	8,8	0.0286	0.0229	0.0180	0.2123	0.0463	0.0317	0.0286	0.0254	0.0254
	10,10	0.0329	0.0306	0.0271	0.2351	0.0477	0.0337	0.0329	0.0309	0.0309
	15,8	0.0311	0.0311	0.0277	0.2617	0.0411	0.0311	0.0311	0.0300	0.0326
	15,15	0.0334	0.0306	0.0269	0.2643	0.0469	0.0306	0.0334	0.0303	0.0303
	20,15	0.0280	0.0263	0.0234	0.2697	0.0374	0.0243	0.0274	0.0266	0.0171
	20,20	0.0366	0.0343	0.0309	0.2786	0.0440	0.0314	0.0366	0.0351	0.0351
	40,30	0.0317	0.0303	0.0286	0.3089	0.0349	0.0249	0.0311	0.0294	0.0197
	80,100	0.0449	0.0440	0.0429	0.3423	0.0483	0.0360	0.0449	0.0443	0.0569
0.15	8,8	0.0411	0.0377	0.0300	0.2540	0.0649	0.0454	0.0411	0.0374	0.0374
	10,10	0.0400	0.0374	0.0320	0.2706	0.0637	0.0411	0.0394	0.0383	0.0383
	15,8	0.0380	0.0380	0.0343	0.3123	0.0543	0.0383	0.0377	0.0366	0.0394
	15,15	0.0426	0.0426	0.0377	0.2957	0.0580	0.0429	0.0426	0.0403	0.0369
	20,15	0.0417	0.0391	0.0343	0.2960	0.0534	0.0394	0.0409	0.0383	0.0383
	20,20	0.0466	0.0443	0.0406	0.3289	0.0589	0.0420	0.0463	0.0451	0.0451
	40,30	0.0554	0.0551	0.0526	0.3966	0.0600	0.0474	0.0549	0.0540	0.0377
	80,100	0.0900	0.0889	0.0869	0.4817	0.0903	0.0757	0.0900	0.0886	0.1077
0.35	8,8	0.0697	0.0649	0.0551	0.3231	0.1009	0.0757	0.0706	0.0671	0.0671
	10,10	0.0740	0.0677	0.0557	0.3617	0.1077	0.0760	0.0737	0.0669	0.0669
	15,8	0.0820	0.0820	0.0720	0.4157	0.1034	0.0817	0.0803	0.0754	0.0669
	15,15	0.0923	0.0880	0.0823	0.4134	0.1086	0.0894	0.0931	0.0877	0.0877
	20,15	0.1023	0.1009	0.0917	0.4740	0.1283	0.0934	0.1017	0.0980	0.0697
	20,20	0.1000	0.0977	0.0914	0.4483	0.1163	0.0940	0.1006	0.0974	0.0974
	40,30	0.1649	0.1623	0.1563	0.6131	0.1654	0.1386	0.1649	0.1614	0.1183
	80,100	0.3006	0.2971	0.2937	0.7646	0.3060	0.2634	0.3017	0.2977	0.3457
0.50	8,8	0.0854	0.0749	0.0591	0.3714	0.1166	0.0949	0.0863	0.0777	0.0777
	10,10	0.1151	0.1063	0.0911	0.4320	0.1314	0.1206	0.1160	0.1097	0.1097
	15,8	0.1143	0.1143	0.1031	0.4940	0.1423	0.1146	0.1146	0.1063	0.0929
	15,15	0.1314	0.1251	0.1140	0.5049	0.1537	0.1257	0.1326	0.1246	0.1246
	20,15	0.1483	0.1463	0.1340	0.5703	0.1654	0.1377	0.1477	0.1409	0.1049
	20,20	0.1589	0.1540	0.1449	0.5463	0.1820	0.1477	0.1589	0.1534	0.1534
	40,30	0.2694	0.2686	0.2629	0.7134	0.2794	0.2440	0.2709	0.2657	0.2100
	80,100	0.5206	0.5169	0.5137	0.8629	0.5260	0.4780	0.5200	0.5177	0.5660
0.75	8,8	0.1546	0.1394	0.1180	0.4577	0.1831	0.1657	0.1551	0.1466	0.1466
	10,10	0.1746	0.1620	0.1386	0.5203	0.2129	0.1774	0.1740	0.1634	0.1634
	15,8	0.2069	0.2069	0.1911	0.6143	0.2397	0.2071	0.2066	0.1960	0.1134
	15,15	0.2374	0.2260	0.2103	0.6157	0.2531	0.2274	0.2369	0.2257	0.2257
	20,15	0.2951	0.2926	0.2763	0.6997	0.3149	0.2826	0.2957	0.2866	0.2349
	20,20	0.3060	0.2980	0.2823	0.6840	0.3234	0.2877	0.3057	0.2974	0.2974
	40,30	0.5240	0.5217	0.5151	0.8509	0.5343	0.4986	0.5249	0.5177	0.4529
	80,100	0.8609	0.8597	0.8569	0.8954	0.8606	0.8351	0.8609	0.8597	0.8869

Note: See Table I

Journal of Mathematical Education in Science and Technology, 2007, pp.412-421.

[9] S. Banik and B.M.G. Kibria, "Comparison of some parametric and nonparametric type one sample confidence intervals for estimating the mean of a positively skewed distribution" Communications in Statistics-Simulation and Computation, 2010, pp.1-29.

[10] B.M.G. Kibria and Banik, S., "Parametric and nonparametric confidence intervals for estimating the differences of means of two skewed populations", Journal of Applied Statistics, 2013, Vol. 40, pp. 2617-2636

[11] B.L. Welch, "The significance of the difference between two means when the population variances are unequal", Biometrika, 1938, pp.350-362.

[12] F.E. Satterthwaite, "An approximation distribution of estimates of variance components", Biometrics Bull,1946, pp.110-114.

[13] N.A.C. Cressie and H.J. Whitford, "How to use the two sample t-test?", Biometrical Journal, 1986, pp.131-148.

[14] J.E. Reed and D.B. Strak, "Huge estimators of location: robust to asymmetry", Computation Methods *Programs in Biomedicine*, 1996 pp.11-17.

TABLE III: ESTIMATED POWERS OF SELECTED TESTS FOR G(0.25,1) AND G(0.75,1) WITH SKEWNESS 4 AND 2.3094 RESPECTIVELY AT THE 5% SIGNIFICANCE LEVEL

c	(n_1,n_2)	t	t_{WS1}	t_{WS2}	ZD1	ZD2	ZD3	t_{JOHN}	t_{MADt1}	t_{MADt2}
0.05	8,8	0.0620	0.0871	0.0734	0.2071	0.0760	0.0980	0.0931	0.0746	0.0746
	10,10	0.0694	0.0846	0.0766	0.2031	0.0820	0.0923	0.0920	0.0734	0.0734
	15,8	0.0774	0.1023	0.1143	0.1917	0.1100	0.1174	0.1206	0.1009	0.0911
	15,15	0.0705	0.0794	0.0740	0.2043	0.0783	0.0783	0.0849	0.0646	0.0646
	20,15	0.0709	0.0831	0.0837	0.1877	0.1094	0.0837	0.0900	0.0723	0.0731
	20,20	0.0710	0.0780	0.0740	0.2040	0.0983	0.0746	0.0809	0.0643	0.0643
	40,30	0.0714	0.0749	0.0749	0.1674	0.1154	0.0689	0.0763	0.0643	0.0671
	80,100	0.0717	0.0757	0.0743	0.2260	0.0849	0.0643	0.0760	0.0620	0.0763
0.15	8,8	0.1011	0.1354	0.1177	0.2497	0.1189	0.1483	0.1420	0.1194	0.1194
	10,10	0.1043	0.1469	0.1263	0.2574	0.1326	0.1566	0.1529	0.1231	0.1231
	15,8	0.1114	0.1680	0.1871	0.2371	0.1660	0.1923	0.1923	0.1703	0.1651
	15,15	0.1209	0.1477	0.1406	0.2489	0.1560	0.1466	0.1526	0.1249	0.1249
	20,15	0.1237	0.1563	0.1563	0.2351	0.1854	0.1569	0.1637	0.1443	0.1480
	20,20	0.1251	0.1626	0.1520	0.2577	0.1743	0.1531	0.1634	0.1351	0.1351
	40,30	0.1714	0.1960	0.1960	0.2089	0.2811	0.1854	0.1997	0.1700	0.1889
	80,100	0.2283	0.3180	0.3154	0.3240	0.3826	0.2889	0.3157	0.2797	0.2537
0.35	8,8	0.1994	0.2446	0.2231	0.3666	0.1909	0.2617	0.2520	0.2211	0.2211
	10,10	0.2077	0.2720	0.2514	0.3809	0.2326	0.2811	0.2774	0.2423	0.2423
	15,8	0.2157	0.2980	0.3211	0.3311	0.2534	0.3257	0.3277	0.2943	0.2886
	15,15	0.2614	0.3380	0.3243	0.3769	0.3003	0.3337	0.3400	0.3014	0.3014
	20,15	0.2726	0.3651	0.3640	0.3394	0.3406	0.3609	0.3711	0.3354	0.3469
	20,20	0.2914	0.3891	0.3760	0.3766	0.3743	0.3749	0.3897	0.3503	0.3503
	40,30	0.4334	0.5200	0.5186	0.2846	0.5567	0.4977	0.5191	0.4826	0.5174
	80,100	0.6709	0.8809	0.8786	0.5494	0.8156	0.8614	0.8783	0.8540	0.8286
0.50	8,8	0.2994	0.3734	0.3423	0.4140	0.2574	0.3977	0.3794	0.3406	0.3406
	10,10	0.3186	0.4151	0.3897	0.4049	0.2983	0.4263	0.4174	0.3774	0.3774
	15,8	0.3354	0.4089	0.4294	0.3497	0.3291	0.4360	0.4354	0.4063	0.4254
	15,15	0.4106	0.5074	0.4926	0.4054	0.4189	0.5040	0.5103	0.4697	0.4697
	20,15	0.4483	0.5386	0.5380	0.3503	0.4514	0.5369	0.5434	0.5103	0.5380
	20,20	0.4897	0.5877	0.5740	0.3971	0.5071	0.5740	0.5831	0.5454	0.5454
	40,30	0.5900	0.7360	0.7354	0.3760	0.6254	0.7151	0.7340	0.7011	0.7426
	80,100	0.8331	0.9931	0.9923	0.6311	0.8350	0.8900	0.9920	0.9894	0.9829
0.75	8,8	0.4969	0.5843	0.5540	0.4643	0.3223	0.6109	0.5909	0.5589	0.5589
	10,10	0.5629	0.6526	0.6360	0.5706	0.3909	0.6691	0.6554	0.6340	0.6340
	15,8	0.5386	0.5989	0.6211	0.5869	0.3443	0.6306	0.6254	0.6000	0.6749
	15,15	0.6703	0.7669	0.7554	0.6177	0.4343	0.7666	0.7646	0.7389	0.7389
	20,15	0.6954	0.7857	0.7866	0.6671	0.4391	0.7883	0.7906	0.7657	0.8026
	20,20	0.7529	0.8494	0.8414	0.6880	0.5640	0.8446	0.8443	0.8263	0.8263
	40,30	0.8514	0.9500	0.9500	0.7091	0.6369	0.9437	0.9480	0.9406	0.9583
	80,100	0.8921	1.0000	1.0000	0.7189	0.8977	0.9523	1.0000	1.0000	1.0000

Note: See Table I

[15] N.J. Johnson, "Modified t-tests and confidence intervals for asymmetrical population" Journal of American Statistical Association, 1978, pp.536-544.

[16] G.P. Rossi, T. Stefano, V. Agostino, C. Martina, G. Lorenzo, F. Stefania, V. Daniele, S. Isabella, C.P. Achille and S. Antonio, "The T-786C and Glu298Asp polymorphisms of the endothelial nitric oxide gene affect the forearm blood flow responses of Caucasian hypertensive patients", Journal of the American College of Cardiology, 2003, pp.938-945.

Diagnostics based Principal Component Analysis for Robust Plane Fitting in Laser Data

Abdul Nurunnabi †, David Belton ‡, Geoff West‡
Department of Spatial Sciences
Curtin University, Perth, Australia
CRC for Spatial Information (CRCSI)
†abdul.nurunnabi@postgrad.curtin.edu.au, ‡{d.belton, g.west}@curtin.edu.au

***Abstract*— Plane fitting and obtaining characteristics (e.g., normal) from the estimated plane are fundamental tasks in many applications in which laser scanner 3D data is used. Unfortunately, laser data are not free from outliers. Principal Component Analysis (PCA) is a popular method for plane fitting, but it is known that PCA is very sensitive to outliers and gives misleading non-robust results. We present a robust plane fitting algorithm based on PCA coupled with an outlier detecting diagnostic statistical approach. In this method, the recently introduced robust scatter matrix is used to calculate robust statistical distance for finding outliers. After excluding outliers, PCA is performed on the outlier free data which is used for fitting planar surfaces and to estimate robust normal and other parameters. Demonstration of the new algorithm through several synthetic and vehicle based laser scanning data show that the proposed method is efficient, and gives robust estimates. Results outperform Least Squares (LS), PCA and are significantly better than the well-known RANSAC in terms of time, accuracy and robustness. This method has great potential for robust segmentation, surface reconstruction, and other point cloud processing tasks.**

***Index Terms*— Feature extraction, outlier, robust normal, point cloud, segmentation, surface reconstruction.**

I. Introduction

Laser scanning for data acquisition in 3D modelling, surface reconstruction and object recognition has drawn a great deal of attention as a useful and competitive technology in many subjects include augmented reality, computer vision, computer graphics, photogrammetry, reverse engineering, remote sensing, simulation and visualization over the years [1]-[4]. Applications in these fields require accurate and robust planar surface fitting. Robust surface fitting is a challenging task because the point cloud data is not free from outliers. Fitting planes and normal estimation is closely related to surface reconstruction. The goodness of normal estimation depends on good quality plane fitting and is necessary for true surface reconstruction, point cloud segmentation, classification and rendering [1], [2], [4].This is because the fastest and more robust surface reconstruction algorithms require points with normals as input, instead of unorganized points [2]. The presence of outliers is a common phenomenon in data analysis. Simply, an outlier is an observation that deviates so much from the majority observations as it seems that it was generated from a different population. The majority of point cloud data is acquired by various measurement processes using a number of instruments (sensors). Usually, the physical limitations of the sensors, boundaries between 3D features, occlusions, multiple reflectance and noise can produce off-surface points that appear to be outliers [5].

Local normal estimation from a fitted local plane is used frequently [6]-[8] in laser scanning data and range image processing. The Least Squares (LS) method and its equivalent Principal Component Analysis (PCA) are the two most popular classical techniques for planar surface fitting [9]. But it is well-known that the methods are much influenced by outliers and fail to reliably fit planes. RANdom SAmple Consensus (RANSAC) is a well-known method frequently used for model parameter estimation [10]. The RANSAC algorithm is not completely free from the effect of outliers and requires more processing time for large datasets. It is shown in this paper that the RANSAC algorithm can misclassify inliers into outliers and vice versa. Diagnostic statistical approaches have been regarded as one of the most effective solutions to the problem of outliers. It put more emphasis on outlier detection rather than the fitting procedure. However, sometimes it is apparent that non-robust diagnostic estimators are not sufficient for outlier detection in multivariate data. Apart from diagnostics, robust statistics is necessary to produce estimators that are less influenced by outliers and even large changes in data structure should not cause an unacceptable breakdown for the estimators [11]. Many robust diagnostic estimators are introduced in the literature [11], [12].The goal of robust diagnostic methods can be described in two ways: (i) identifying outliers (ii) providing an analysis that result in greatly reduced sensitivity to outliers.

In this paper, we propose an algorithm for robust planar surface fitting and parameter estimation. Robust versions of PCA often portray an accurate figure of the underlying data structure, but they do not always identify outliers that may be significant in their own right for the analysis. To get more effective results, along with PCA, it is important to have a complementary outlier detection method. In the proposed algorithm, we are able to find outliers and to get robust estimates at the same time. In addition, the new method is compared with classical (LS and PCA) methods and

RANSAC. The proposed method is compared with respect to size of the data, outlier percentage, surface roughness, data classification accuracy, speed of computation, edge detection and segmentation.

The remainder of this paper is organized as follows: Section II briefly discusses the related methods. In Section III, we propose our technique for fitting planes, robust normal and for other local saliency feature estimation. Section IV contains results, and analyses the performance of the proposed technique through comparison with the other methods using simulated and real (laser scanning) datasets.

II. EXISTING POPULAR METHODS

This section includes brief discussions about the basics of some popular methods used later for comparison.

A. Least Squares

Minimization of the sum of the squared residuals known as the Least Squares (LS) principle is the most common parameter estimation method. For a sample of 3D data points $\{p_i(x_i, y_i\ z_i)\}$, a plane equation can be written as:

$$ax + by + cz + d = 0\,, \tag{1}$$

where a, b, c and d are the parameters. The classical LS method works as in regression analysis and makes a functional relation among the three variables as: $z = f(x, y)$. The LS minimizes the sum of the squared residuals:

$$\min \textstyle\sum_{i=1}^{n} r_i^2 = \min \sum_{i=1}^{n} (z_i - \hat{z}_i)^2\,, \tag{2}$$

where r_i is the distance (residual). Since the ith residual is a vertical distance and considered only in one (vertical) direction (z here), many authors criticize it for its one way minimization. To overcome the bias in one direction, one technique is fitting the plane in Eq. 1 using the eigenvalue method under the constraint:

$$a^2 + b^2 + c^2 + d^2 = 1 \tag{3}$$

i.e. the optimal estimation of the plane to the plane normal direction [8].

B. Principal Component Analysis

PCA is a well-known statistical technique for data visualization and dimension reduction, which is typically used to identify a small set of mutually orthogonal variables that can explain the underlying variance pattern of a dataset. The Principal Components (PCs), are the linear combination of the original variables that rank the variability in the data through the variances, and produce the corresponding directions. Every PC describes a part of the data variance not explained by the others. The first PC corresponds to the direction in which the projected observations have the largest variance, then the second PC does the same for the second largest variance and so on. The main philosophy for PCA is: using only the top ranked PCs enables a reduced representation of the data and visualizes the underlying data pattern in the lower dimensions. In the case of plane fitting, the first two PCs explain most of the variability and are sufficient to represent a 2D linear approximation, so the fitted plane is the best fit to the data. PCA works on the mean centred covariance matrix. It finds the required number of PCs by performing Singular Value Decomposition (SVD) on that covariance matrix. Usually, the 3rd PC is orthogonal to the first two PCs and defines the normal to the fitted plane. Since the third PC expresses the least amount of variation, it can be used as the estimate of the coefficients of the fitted plane [6]. Many authors use the PCA approach in many ways [7]-[9] as an equivalent to the LS method. Although PCA is very popular, it is shown that the results are affected by outliers and can give misleading estimates.

C. RANSAC

Fischler and Bolles [10] introduce RANdom SAmple Consensus (RANSAC). It has been recognized as a very competitive technique for shape estimation and to estimate the parameters of a model. It is conceptually simple and can give robust results for the data even when contaminated with more than 50% of outliers [13]. It works as an iterative approach consisting of two steps: hypothesize and verification. First, a minimal subset (e.g. three points for a plane) is randomly selected and the required parameters are estimated by using a LS objective function based on the minimal subset. Second, in the verification phase, the estimates are tested against all the data points to determine how many of them are consistent with the estimates. The iteration is continued until the maximum inlier set is found. RANSAC estimates parameters finally based on the maximum set of inliers. Many authors use RANSAC for planar surface fitting [6], [13]. RANSAC can be sensitive to the choice of the error limit and the presence of multiple model instances. The requirement for large amounts of memory is its main drawback [13].

III. PROPOSED ALGORITHM

In this section, we propose an algorithm that combines a statistical outlier diagnostic approach and the classical PCA. This section also outlines the basic principles and the variants of the distance based outlier detection approach in multivariate data which are used in the proposed technique.

Multivariate outliers are complex in nature and proper detection of them is an important and challenging problem, especially in the presence of cluster/multiple outliers. A well-known multivariate method for outlier detection is the distance based approach, in which the shape (covariance matrix) of the data is considered together with the distance of an observation from the centre of the data. Mahalanobis Distance (MD) is the most popular classical method that considers both location (centre) and shape of the data. It is defined as:

$$MD_i = \sqrt{(x_i - c)^T \Sigma^{-1} (x_i - c)}, \quad i = 1,2,\ldots,n \tag{4}$$

where x is an m dimensional variable, c is the arithmetic mean and $\sum$ is the covariance matrix of the sample data. Rousseeuw and Driessen [14] state: "Although it is still quite easy to detect a single outlier by means of MD, this approach no longer

suffices for multiple outliers because of the well-known masking effect". Masking occurs when an outlying observation appears as an inlier because of the presence of other outliers, usually from an adjacent group of points. The MD is not robust because of the outlier sensitivity of the mean and covariance matrix [11].

A. Robust Distance

Hubert et al. [12] show the necessity of using robust estimators of multivariate location and scatter. Many robust estimators have been introduced in the literature [12], [14], [15] for getting robust location and scatter. Insertion of M-estimators for centre and scatter is a marked extension but the breakdown point [11] of the M-estimator is not high. A change occurs when the Minimum Volume Ellipsoid (MVE) has been introduced [15]. MVE is an ellipsoid of the smallest volume of h points. The MCD, another estimator has been introduced by Rousseeuw [15] that finds a subset of h ($(n/2) \le h \le n$) observations with the smallest determinant of the covariance matrix. The MCD estimator of covariance matrix has better statistical efficiency and accuracy than MVE [14].The Robust Distance (RD) based on the MCD estimators is defined as:

$$RD_i = \sqrt{(x_i - c_{MCD})^T \Sigma_{MCD}^{-1} (x_i - c_{MCD})}, \quad i = 1,2,\ldots,n \quad (5)$$

where c_{MCD} and Σ_{MCD} are the MCD estimates of location and shape of the h points respectively. Rousseeuw and van Zomeren [16] show that RD follows a Chi-square (χ^2) distribution with m degrees of freedom, and the observations that exceed $\sqrt{(\chi^2_{m,0.975})}$ are identified as outliers. Unfortunately, finding the h subset for the MCD is not easy and takes much time. After introducing the fast-MCD [14], the MCD based RD becomes popular because of its speed. Interested readers are referred to [14] for detail about fast-MCD.

B. Deterministic MCD

Recently, Hubert et al. [17] propose a deterministic algorithm for robust location and scatter, denoted as DetMCD. The authors claim that fast-MCD is affine equivariant but not permutation invariant. The DetMCD is permutation invariant and even faster than fast-MCD. It uses the same iteration steps but does not use random subsets. Rather it starts from only a few well-chosen initial estimates. The DetMCD uses parts of both of the fast-MCD [14] and Orthogonalized Gnanadesikan-Kattenring (OGK) estimator [18]. We use the DetMCD to get a robust version of MD by plugging the robust estimators of location and scale based on DetMCD in place of c and $\sum$ in Eq. 4. The DetMCD based robust distance denoted as DRD, can be defined as:

$$DRD_i = \sqrt{(x_i - c_{DMCD})^T \Sigma_{DMCD}^{-1} (x_i - c_{DMCD})}, \; i = 1,2,\ldots,n \quad (6)$$

C. Implementation

In terms of point cloud processing, for local surface fitting, we need to find the local neighbourhood for every point in the local surface. We use the well-known k-Nearest Neighbourhood (k-NN) based algorithm for finding local neighbours. We choose the k-NN approach to avoid the problem of point density variation in the data. Note that if we want to fit a plane for a full dataset, we do not need neighbourhood searching. Our algorithm works in two stages. In the first stage, we calculate robust distance based on Eq. 6 and find outliers using the DRD values. We use the same cut-off value for DRD as for the MD and RD. Excluding outliers reduces outlier effects within the local neighbourhood and makes the data more homogeneous. Second, classical PCA is performed on the covariance matrix without the outlying cases. Now we get the required PCs. Projecting the data points onto the first two PCs, we get the best fitted plane. The 3rd PC is the required normal, which is also used for the estimated parameters. We name the method DRD-PCA. The algorithm consists of the followings steps.

i. For local planar surface fitting, find the k nearest neighbourhood for every point in the data. Otherwise, consider the whole dataset as one sample for which we want to fit a plane.
ii. Calculate the DRD values using Eq. 6 for the points.
iii. Classify regular observations and outliers based on the DRD values:

$$DRD_i = \begin{cases} < \chi^2_{m,0.975}, & i \in R \\ \ge \chi^2_{m,0.975} & i \in O, \end{cases} \quad i = 1,\ldots,k \quad (7)$$

where R and O are the sets of Regular points and Outliers.

iv. Perform PCA for the regular observations from Step (iii).
v. Find and rank the first three PCs with the highest eigenvalues from PCA in Step (iv).
vi. Project observations onto the two first PCs to get the fitted plane. The 3rd PC is the plane normal.

In a similar fashion, we also perform MD and RD based PCA to generate two alternative methods for plane fitting. They are dubbed as MD-PCA and RD-PCA respectively.

IV. Experiments and Evaluation

In this section, the proposed method is demonstrated though several sets of artificial and vehicle based laser scanning (mobile mapping) real data. We compare our method with existing LS, PCA and RANSAC methods. At the same time we show the comparative performance of the results from MD-PCA and RD-PCA. The performance of the plane estimates are in the context of applications of surface edge detection and point cloud segmentation is shown by laser scanner data in Sub-section B of this section.

To show the performance, we fit the planar surface by different methods, and find normal and eigenvalue characteristics. As the measures of performance, we calculate (i) the bias angle θ between the planes fitted to the data with and without outliers, which can be defined as [8]:

$$\theta = \arccos \mid \hat{n}_1^T . \hat{n}_2 \mid \quad (8)$$

where $\hat{n}_1$ and $\hat{n}_2$ are the two unit normals from the fitted planes with and without outliers respectively, (ii) the variation along the plane normal and (iii) the surface variation [7]. Variation along the normal is defined by the corresponding least eigenvalue λ_0, and the surface variation along the direction of the corresponding eigenvectors at the point p_i known as curvature is defined as:

$$\sigma(p_i) = \frac{\lambda_0}{\lambda_0 + \lambda_1 + \lambda_2}, \qquad \lambda_0 \le \lambda_1 \le \lambda_2 \qquad (9)$$

where λ_i is the ith eigenvalue. The λ_0 and σ_p values along with the normals are evaluated in applications in point cloud processing (surface edge detection and segmentation).

A. Artificial Data

The artificial datasets are generated randomly from two sets of 3D multivariate Gaussian distributions. The first set is for regular observations with means (2, 2, 2) and variances (6, 6, 0.02) and the second set is for outlying cases, they have means (7, 9, 11) and variances (7, 4, 0.5).We simulate the datasets for different sample sizes and outlier percentages.

1) *Bias Angle Evaluation:* We compute and measure the performance based on bias angle θ in different perspectives. Table I contains the average bias angles (in radians) based on 1000 runs for sample sizes 50 and 100, and outlier percentages 5, 10, 15, 20 and 25. It shows large gaps between robust (RANSAC, RD-PCA and DRD-PCA) and non-robust (LS, PCA and MD-PCA) methods. We see little difference between LS and PCA bias angles, although PCA is more consistent than LS, and MD-PCA performs reasonable with only lower outlier contamination (5%). RANSAC produces smaller bias angles than the non-robust methods but always gives larger bias angles than RD-PCA and DRD-PCA. For example, for sample size 100 with 10% outliers, RANSAC has average θ=0.0146, which is 3.6 times larger than the value of DRD-PCA (0.0041). RD-PCA and DRD-PCA generally give almost similar results to each other.

TABLE I. AVERAGE BIAS ANGLES FOR DIFFERENT SAMPLE SIZES AND OUTLIER PERCENTAGES

Sample Size	Methods	Outlier Percentage				
		5	*10*	*15*	*20*	*25*
50	LS	0.4522	0.5575	0.8399	0.9882	1.0162
	PCA	0.5366	0.6275	0.7044	0.7441	0.7593
	MD-PCA	0.0053	0.3690	0.7036	0.7477	0.7701
	RANSAC	0.0165	0.0180	0.0216	0.0225	0.0272
	RD-PCA	0.0075	0.0097	0.0133	0.0131	0.0135
	DRD-PCA	0.0070	0.0073	0.0098	0.0093	0.0112
100	LS	0.3116	0.5813	0.7364	0.8588	0.9422
	PCA	0.4684	0.6185	0.6720	0.7021	0.7239
	MD-PCA	0.0031	0.3274	0.6613	0.7041	0.7335
	RANSAC	0.0162	0.0146	0.0188	0.0176	0.0197
	RD-PCA	0.0035	0.0052	0.0061	0.0056	0.0061
	DRD-PCA	0.0034	0.0041	0.0052	0.0048	0.0054

Fig. 1(a) presents boxplots of the bias angles from 1000 runs of datasets of 100 observations with 15% outliers. The plots show that non-robust techniques (LS, PCA and MD-PCA) have larger bias angles than the three robust methods and show more variation. Fig. 1(b), shows that RD-PCA and DRD-PCA are more robust and produce less outlying results than the RANSAC.

We know that surface thickness/roughness can influence surface fitting methods. To see the effect of roughness on the bias angles, we change the variance along the z (elevation) axis. We simulate 1000 datasets of 100 points with 15% outliers. The z variances for regular observations are changed to 0.001, 0.01, 0.02, 0.05 and 0.1, and the rest of the parameters remain the same as for the previous experiment.

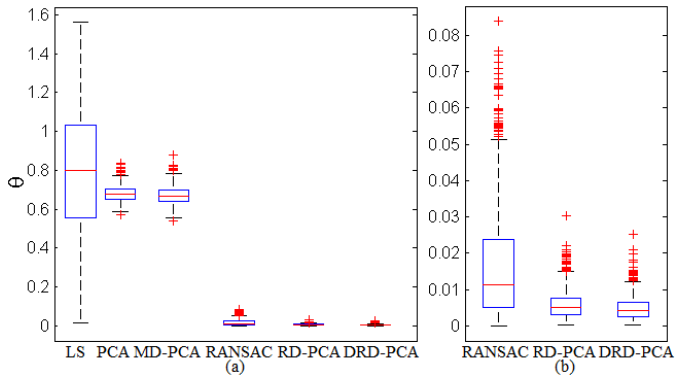


Fig. 1. Box plots for bias angle (a) all methods (b) robust methods.

The average bias angles are calculated for all 1000 datasets, and the results of average bias angles are shown in Fig. 2. In Fig. 2(a), we see larger average θ values for non-robust methods. Fig. 2(b) clearly shows that RANSAC gets worse compared to the others (RD-PCA and DRD-PCA) with increasing z-variation.

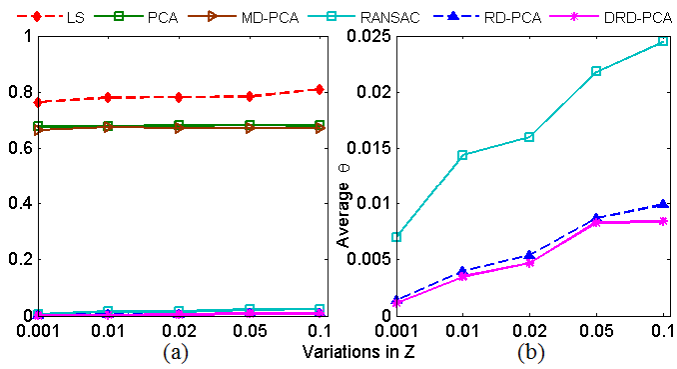


Fig. 2. Effects of variations in Z on average bias angle; (a) all methods (b) only robust methods.

2) *Diagnostics Accuracy:* To see the accuracy of inlier and outlier identification, we calculate the Correct Inlier Identification Rate (CIIR) and the Masking Rate (MR; the rate of outliers identified as inliers), which is a misclassification rate. We generate artificial datasets as for previous experiments of 100 points with 15% outliers. We run the

algorithm 1000 times and classify the outliers and inliers and take the average results. Results in Table II show that DRD has the highest CIIR without any masking and MD has a huge masking effect. In comparison, RANSAC always identifies inliers less than 50% that means it is affected by swamping, i.e. it shows inliers as outliers. We can say that the RANSAC planes are ill-fitted planes because of lower inlier identification.

TABLE II. DATA POINTS CLASSIFICATION

Methods		Outlier Percentage				
		5	*10*	*15*	*20*	*25*
MD	**CIIR**	99.09	99.06	99.15	98.99	98.97
	MR	0	29	84.73	93.55	96.80
RANSAC	**CIIR**	43.03	42.34	43.18	43.04	45.90
	MR	0	0	0	0	0
RD	**CIIR**	93.67	94.80	95.282	95.9	96.76
	MR	0	0	0	0	0
DRD	**CIIR**	94.83	95.64	96.33	96.24	97.04
	MR	0	0	0	0	0

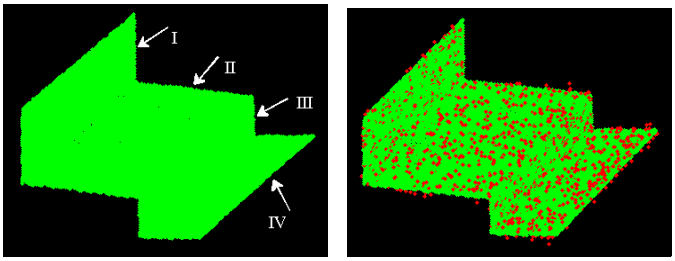
3) Computation Time: Table III shows the time (in seconds) for on average plane fitting. Times are calculated by using the MATLAB@ profile function. The data are generated similar to the previous experiments. Results are for various sample sizes (with 15% outliers and z-variance 0.02) based on 1000 runs. We see that although DRD-PCA takes more time than the non-robust methods, its time is significantly less than RANSAC and RD-PCA. For a data size of 1,000, DRD-PCA takes only 0.1385 sec. which is 2.34 and 7.64 times less than RANSAC (0.3246) and RD-PCA (1.0581) respectively.

TABLE III. MEAN COMPUTATION TIME

Sample size	LS	PCA	MD-PCA	RANSAC	RD-PCA	DRD-PCA
100	0.0001	0.0004	0.0013	0.2094	0.8306	0.0322
500	0.0002	0.0005	0.0013	0.2604	0.9624	0.0595
1000	0.0001	0.0006	0.0014	0.3246	1.0581	0.1385
10000	0.0012	0.0016	0.0031	1.6529	1.1519	0.4437
20000	0.0024	0.0029	0.0049	2.2606	1.1910	0.6725

B. Laser Scanner 3D Point Cloud Data

To see the quality of the estimates (normal, λ_0, and curvature σ_p) from the fitted planes, we use the estimates for edge detection and segmentation on point clouds. We use two recently proposed edge detection [19] and segmentation [20] algorithms. We take a vehicle based (mobile mapping) laser scanning real 3D point cloud dataset (Fig. 3(a)) with 13,259 points. The dataset contains 4 separate planar surfaces (I to IV). To make sure the dataset is noisy, we add 15% multivariate Gaussian noise with mean and standard deviation (0, 0.03) in every direction (Fig. 3(b)).



(a) (b)

Fig. 3. (a) Laser scanning data (b) data with 15% noise.

1) *Surface Edge Detection:* For edge detection we follow the rule in [19]. We calculate λ_0 values for every point in the point cloud in Fig. 3(b). The ith point is considered as an edge point if its corresponding least eigenvalue λ_0 is:

$$\lambda_0 > mean(\lambda_0) + 1\times \textit{ standard deviation } (\lambda_0). \tag{10}$$

The results in Fig. 4 show that the non-robust methods misclassify the surface and edge points. They do not identify edge points correctly. Even RANSAC fails to identify edge points properly. RD-PCA can identify edge points well and DRD-PCA identifies them more accurately than any other methods.

2) *Segmentation:* Segmentation is the labelling of the most similar surface points into a number of separate surfaces. The region growing based segmentation algorithm [20] begins by searching for a seed point that has the least curvature value. It grows regions using local surface proximity and intrinsic criteria based on local neighbourhood (NP_i) of the points. The algorithm puts the jth observation in the same region as the ith seed point if they have (i) Orthogonal Distance (OD; distance between the jth point and the best-fit-plane of the ith seed point), (ii) Euclidian Distance (ED), and (iii) the angle difference between them less than pre-assigned thresholds. Readers are referred to [20] for more about the algorithm. We set the parameters: neighbourhood size k=30, θ^{th}=10° and minimum region size R_{min}=10.

Fig. 5 shows that LS and PCA fail to segment planes I and III. MD-PCA and RANSAC creates over segmentation for planes III and I respectively. RD-PCA has slight over segmentation of plane I and the DRD-PCA gives the best segmentation results without any over or under segmentation.

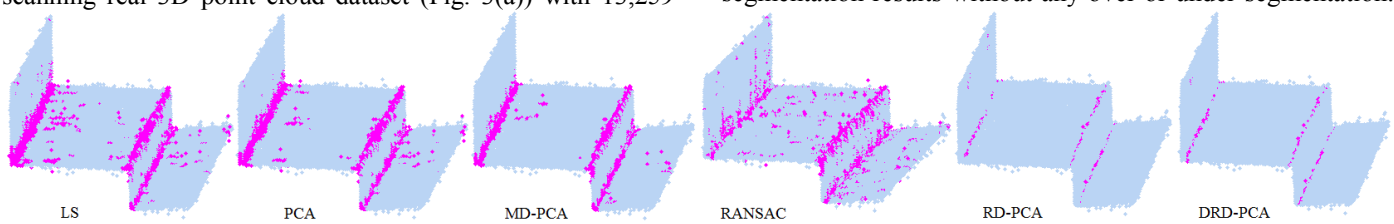


Fig. 4. Surface edge detection.

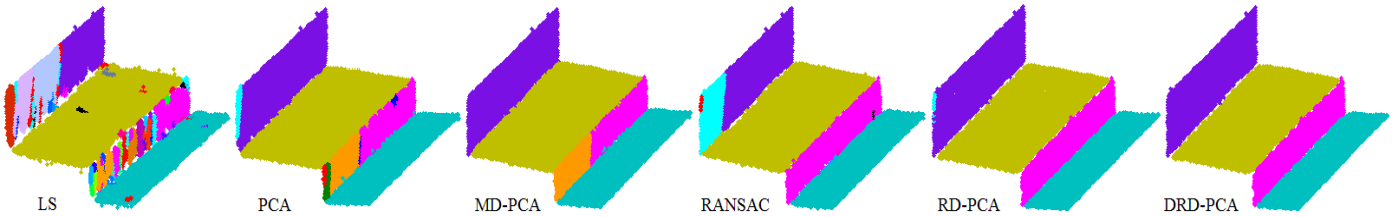


Fig. 5. Segmentation.

V. CONCLUSION

This paper proposes a robust diagnostic approach based PCA algorithm for fitting robust plane in 3D laser data. Experimental results based on several artificial and real laser scanning datasets show that the proposed diagnostic technique outperforms classical LS and PCA methods. It also performs better than RANSAC and is significantly faster. Application of the estimated normal and other local saliency features for edge detection and segmentation prove their better accuracy, efficiency and robustness. Similar to other popular robust techniques, the proposed method breaks down for more than 50% outliers. The diagnostics based PCA technique has great potential for surface reconstruction, object recognition and 3D modelling.

ACKNOWLEDGMENT

This study has been carried out as a PhD research supported by a Curtin University International Postgraduate Research Scholarship (IPRS). The work has been supported by the Cooperative Research Centre for Spatial Information (CRCSI), whose activities are funded by the Australian Commonwealth's Cooperative Research Centres Programme. We are also thankful to McMullen Nolan and Partners Surveyors for the real point cloud datasets.

REFERENCES

[1] B. Li, R. Schnabel, R. Klein, Z. Cheng, G. Dang, and S. Jin, "Robust normal estimation for point clouds with sharp features," Computers and Graphics, vol. 34, pp. 94–106, 2010.

[2] M. Yoon, Y. Lee, S. Lee, I. Ivrissimtzis, and H-P. Seidel, "Surface and normal ensembles for surface reconstruction," Computer Aided Design, vol. 39, pp. 408–420, 2007.

[3] B. Yang, L. Fang, and J. Li, "Semi-automated extraction and delineation of 3D roads of street scenes from mobile laser scanning point clouds," ISPRS Journal of Photogrammetry and Remote Sensing, vol. 79, pp. 80–93, 2013.

[4] D. Tamal, L. Gang, and J. Sun, "Normal estimation for point cloud: a comparison study for a voronoi based method," Proc. of the Eurographics Symposium on Point-Based Graphics, pp. 39–46, 2005.

[5] S. Sotoodeh, "Outlier detection in laser scanner point clouds," In IAPRS, Dresden, vol. XXXVI/5, pp. 297–301, 2006.

[6] H. Hoppe, T. De Rose, and T. Duchamp, "Surface reconstruction from unorganized points," Proc. of ACM SIGGRAPH, vol. 26(2), pp. 71–78, 1992.

[7] M. Pauly, M. Gross, and L. P. Kobbelt, "Efficient simplification of point sample surface," Proc. of the Conference on Visualization, Washington, D.C., pp. 163–170, 2002.

[8] C. Wang, H. Tanahashi, and H. Hirayu, "Comparison of local plane fitting methods for range data," IEEE Computer Society Conference on CVPR, Kauai, vol. 1, pp. 663–669, 2001.

[9] K. Klasing, D. Althoff, D. Wollherr, and M. Buss, "Comparison of surface normal estimation methods for range sensing applications," IEEE International Conference on Robotics and Automation, pp. 3206–3211, 2009.

[10] M. A. Fischler, and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," Communications of the ACM, vol. 24, pp. 381–395, 1981.

[11] F. Hampel, E. Ronchetti, P. J. Rousseeuw, and W. Stahel, Robust Statistics: The Approach Based on Influence Functions. New York: John Wiley, 1986.

[12] M. Hubert, P. J. Rousseeuw, and S. V. Aelst, "High-breakdown robust multivariate methods," Statistical Science, vol. 23(1), pp. 92–119, 2008.

[13] R. Schnabel, R. Wahl, and R. Klein, "Efficient RANSAC for point-cloud shape detection," The Eurographics Association and Blackwell Publishing, pp.1–12, 2007.

[14] P. J. Rousseeuw, and K. V. Driessen, "A fast algorithm for the minimum covariance determinant estimator," Technometrics, vol. 41(3), pp. 212–223, 1999.

[15] P. J. Rousseeuw, "Least median of squares regression," Journal of the American Statistical Association, vol. 79(388), pp. 871–880, 1984.

[16] P. J. Rousseeuw, and B. C. van Zomeren, "Unmasking multivariate outliers and leverage points," Journal of the American Statistical Association, vol. 85(411), pp. 633–639, 1990.

[17] M. Hubert, P. J. Rousseeuw, and T. Verdonck, "A deterministic algorithm for robust location and scatter," Journal of Computational and Graphical Statistics, vol. 21(3), pp. 618–637, 2012.

[18] R. A. Maronna, and R. H. Zamar, "Robust estimation of location and dispersion for high-dimensional datasets," Technometrics, vol. 44(4), pp. 307–317, 2002.

[19] A. Nurunnabi, D. Belton, and G. West, "Robust segmentation for multiple planar surface extraction in laser scanning 3D point cloud data," IEEE 21st International Conference on Pattern Recognition (ICPR), Tsukuba, Japan, pp. 1367–1370, 2012.

[20] A. Nurunnabi, D. Belton, and G. West, "Robust segmentation in laser scanning 3D point cloud data," Proc. of Digital Image Computing: Techniques and Applications (DICTA), Fremantle, Australia, 3–5 December, 2012.

Optimizing Software Design Migration from Structured Programming to Object Oriented Paradigm

Saeed Siddik
Institute of Information Technology
University of Dhaka, Bangladesh
siddik.saeed@gmail.com

Alim Ul Gias
Institute of Information Technology
University of Dhaka, Bangladesh
alimulgias@gmail.com

Shah Mostafa Khaled
Institute of Information Technology
University of Dhaka, Bangladesh
khaled@univdhaka.edu

***Abstract*—Several industries are using legacy softwares, developed with Structured Programming (SP) approach, that should be migrated to Object Oriented Paradigm (OOP) for ensuring better software quality parameters like modularity, manageability and extendability. Automating SP to OOP migration is pivotal as it could reduce time that take in the manual process. Given this potential benefit, the issue is yet to be addressed by researchers. This paper addresses the scenario by modeling this problem as a graph clustering problem where SP functions and function calls are vertices and edges respectively. The challenge evolving the problem is to find optimized clusters from graphs. To aid this problem, certain heuristic algorithms based on Monte Carlo and Greedy approaches are being developed. The proposed algorithms have been tested against a collection of real and synthetic data. The numerical results show that greedy algorithms are faster and produced better results than the average performance of Monte Carlo based approaches.**

Keywords. Legacy Code, Software Design, Call Graph, DSM, Graph Clustering

I. Introduction

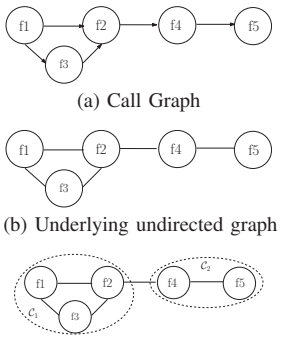
A good number of software are still being used in industries which were developed decades earlier using Structured Programming (SP) approach. These software often go out of support and termed as *Legacy Software*. The source code of such software is commonly referred as *Legacy Code* which becomes unmanageable when it grows too large [1]. Large and complex programs are easier to maintain if developed with Object Oriented Paradigm (OOP) that offers better re-usability, modularity, manageability and extendability by means of encapsulation, inheritance and polymorphism.

In case of bug solving or feature enhancement and integration, legacy codes demand much more time than OOP ones. Such difficulties have their consequential business impacts and thus many industries, dependent on legacy softwares, are facing the issue of business sustainability. Possible way outs could be re-designing the whole product from the scratch or manual SP to OOP design migration which could be error-prone and time consuming. This research intends to propose an approach for automatic SP to OOP design migration.

To analyze the problem we formulate the scenario in terms of a graph where each *function* written in SP code is represented as a vertex, and a *function call* is represented as a directed edge. This gives us a connected directed graph referred as *Call Graph* [2], [3]. The adjacency matrix representation of the call graph produces a 0-1 matrix. The matrix is known as Design Structure Matrix or Dependency Structure Matrix (DSM) [4] which is widely used in industrial engineering for a good variety of applications.

TABLE I. Example of a SP legacy code, its Call Graph and Optimal Clustering

```
function5(....)
{
    ...
}

function4(....)
{
    function5();
    ...
}

function2(....)
{
    function4();
    ...
}

function3(....)
{
    function2();
    ...

}

function1(....)
{
    function2();
    function3();
}
```



(a) Call Graph

(b) Underlying undirected graph

(c) Optimal Clustering

The hypothesis that the methods of a same *class* will call each other more frequently than the number of calls between methods of different classes has been used on the call graph. This hypothesis is based on *encapsulation* [5] that says attributes and *methods* of same *class* or *interface* are more interrelated than attributes and methods of different classes or interfaces [6]. The call graph has an underlying undirected graph which is used to search for vertex clustering having objectives to maximize number of intra-cluster edges and minimize number of inter-cluster edges.

An optimal clustering of the call graph produces number of non-overlapping vertex clusters that represent groups of closely related functions in the SP code. This is used as the clue for migrating the underlying SP design to OOP, where each cluster in the call graph represents a class or interface. The migration is simple as calls between methods of the same class do not change in the new design. However, calls between methods of two different clusters has to be re-written incorporating the issue of making call through proper interfaces between classes.

Table I presents a sample SP legacy code, its corresponding call graph, underlying undirected graph of the call graph and an optimal clustering in of the undirected graph.

A mathematical definition of the optimal cluster finding problem has been formulated in terms of an Integer Program (IP)[7]. It has been realized that the solution to the IP is computationally hard, and therefore different heuristics are applied to get a near optimal solution. A number of randomized and greedy heuristic algorithms has been proposed to assist optimal clustering. The objective function of IP along with conventional Clustering Co-efficient (CC) [8] and Characteristics Path Length(CPL) [9] are used to assess quality of clustering.

The assessment involved 3 data instances, 2 of which were collected from real software and the other is synthetically generated. The results show that greedy huristics were faster and produced better results than the average performance of Monte Carlo based approaches. Moreover, it has been observed that proposed greedy algorithms are 12.8% better in terms of average Clustering Co-efficient (Equation 1), 31.02% better in terms of average Characteristics Path Length (Equation 2)and 25.73% better in terms of average Kal-index (Equation 3) than Monte Carlo algorithms.

Rest of the paper is organized as follows: Section II reviews the background of the problem and the relevant work available in existing literature. A mathematical formulation of the problem is presented in Section III. Section IV presents five variations of our Monte Carlo based heuristics, and three variations greedy algorithms. Numerical results and analysis has been presented in Section V. Section VI concludes the paper with future research direction.

II. Background

Automatic SP to OOP design migration has been rarely addressed as a direct research problem in the existing literature. Although plenty of work has been done with Graph Clustering [10] and DSM [4], existing graph clustering methods mainly focus on the Euclidean distance, but largely ignore vertex connectivity where every distances between two vertices's are same. On the other hand, the researches involving DSM focuses on searching function calls that can be triggered parallely.

Automatic migration from code to design was introduced in [11]. The work converts a COBOL program to a Object Oriented design document. Maqbool et al. reviewed hierarchical clustering research in the context of software architecture recovery and modularization [12]. Moreover, they analyzed the clustering process of multiple clustering algorithms using multiple criteria and showed how arbitrary decisions taken by these algorithms affect the quality of the clusters. In 2011 Dineshkumar et al. presented an empirical approach to migrate from Structured Programming Code to Object Oriented Design [13]. Their work introduced a new technique for code to design migration which creates agglomerative cluster using Jaccard distance matrices.

Zhou et al. proposed a graph clustering algorithm named SA-Cluster based on similar attribute using unified distance measure [14]. This method partitions a large graph associated with attributes into *k*-clusters so that each cluster contains a densely connected sub-graph with homogeneous attribute values. Aggarwal et al. proposed an algorithm to find appropriate sets of clusters and dimensions using medoid technique based on Euclidean coordinate points and preceded by feature selection [15].

Bezdek et al. reviewed two clustering algorithms (hard c-means [16] and single linkage) and three indexes of crisp cluster validity (Huberts statistics, the Davies-Bouldin index, and Dunns index) [17]. Their work illustrates two deficiencies of Dunns index [18] that make it overly sensitive to noisy clusters. They proposed several generalizations of those deficiencies which are not as brittle to outliers in the clusters. Definitions regarding cluster in a graph and measures of cluster quality were reviewed in [10]. This work also presented global algorithms for clustering the entire vertex set of an input graph and discussed the task of identifying a cluster for a specific seed vertex by local computation.

Hossain et al. presented an analytical report on design structure of open source scientic computing software [4]. They used a number of architectural complexity metrics and DSM technique to analyze the design structure. Their analysis involved Automatic Differentiation (AD), Linear Programming (LP) and Mixed Integer Programming (MIP). They have used DSM to present functions that are explicitly implemented in the software under consideration (denoted as user function) and functions that are part of software libraries. Those DSM qualities are measured by characteristic path length, clustering co-efficient, nodal degree, strongly connected components, propagation cost, etc. [9].

Review of existing literature show that none of the work directly proposed a method that focus on converting SP code to detailed object oriented design. We propose a new approach to migrate SP code to OOP design by means of optimal clustering form a call graph. To measure the quality of the clusters a new metric discussed in Section III is proposed. Two existing metrics CC and CPL were also used to measure the quality of clustering.

CC is a measure of degree to which nodes in a graph tend to cluster together [8], [9]. In this paper, local clustering coefficient is used to measure CC (Ψ) index. The local clustering coefficient of a vertex (node) in a graph quantifies how close its neighbors are to being a *complete graph*. Suppose, a graph $G = (V, E)$. An edge $e_{ij} \in E$ connects vertex $v_i \in V$ with vertex $v_j \in V$. The neighborhood N_i for a vertex v_i is defined as its immediately connected neighbors: $N_i = \{v_j : e_{ij} \in E \cap e_{ij} \in E\}$. CC Ψ of an undirected graph is defined as-

$$\Psi = \frac{1}{N}\sum_{i=1}^{N}\Psi_i \quad where \quad \Psi_i = \frac{2\left|\{e_{ij} : v_j, v_k \in N_i, e_{jk} \in E\}\right|}{K_i(K_i - 1)} \tag{1}$$

In Equation (1) Ψ_i denotes the CC (Ψ) of node i and k_i is number of nodes connected to node i, and n_i is actual number of edges within k_i adjacent nodes.

CPL is the distance between pairs of vertices in a connected undirected graph cluster [9]. Let $d(v_i, v_j)$ denote the shortest distance between nodes v_i and v_j, where $\{v_1, v_2\} \in V$ in an unweighed undirected graph G . If $v_1 = v_2$ or v_2 cannot be reached from v_1 then $d(v_i, v_j) = 0$, otherwise $d(v_i, v_j) \geq 1$. Based on these definitions, CPL (χ) of an undirected graph can defined as-

$$\chi = \frac{1}{N(N-1))} \cdot \sum_{i \neq j} d(v_i, v_j) \tag{2}$$

III. Mathematical Formulation of the Problem

Let $G(V, E)$ be the underlying undirected graph of a call graph. V and E be the set of vertices and edges respectively, with $n = |V|, m = |E|$. We define variables x_e and y_e corresponding to each edge $e \in E$, C_v corresponding to each vertex $v \in V$. G may have at most n clusters, where each vertex in the graph will be in a distinct cluster. A vertex, therefore, may potentially be in one of n clusters. The variables are defined as follows:

$$x_e \in \begin{cases} 1 & \text{if } e \text{ is an intra-cluster edge} \\ 0 & \text{otherwise} \end{cases}$$

$$y_e \in \begin{cases} 1 & \text{if } e \text{ is an inter-cluster edge} \\ 0 & \text{otherwise} \end{cases}$$

$$z_{kl} \in \begin{cases} 1 & \text{if vertex } k \text{ belongs to cluster } l \\ 0 & \text{otherwise} \end{cases}$$

$$\mathcal{C}_j \in \begin{cases} 1 & \text{if vertex } j \text{ is the head of a cluster} \\ 0 & \text{otherwise} \end{cases}$$

The problem of maximizing intra-cluster edges, minimizing inter-cluster edges, and maximizing the number of clusters can be formulated as follows:

$$Max \sum_i x_i - \sum_i y_i + \sum_j \mid \mathcal{C}_j \mid, \quad \forall_{i=1,..,m \ and \ j=1,..,n} \tag{3}$$

Subject to:

$$x_i + y_i = 1 \quad \forall_{i=1,2,...,m} \tag{4}$$

$$\sum_{l=1}^{n} z_{kl} = 1 \quad \forall_{k=1,2,...,n} \tag{5}$$

$$\sum_{l=1}^{n} z_{k_a l} \cdot z_{k_b l} = x_{(a,b)} \quad \forall_{(a,b)\in E} \tag{6}$$

$$C_l = \bigcup_k z_{kl} \quad \forall_{k=1,2,...,n} \tag{7}$$

The objective in Equation (3) maximizes the number of intra-cluster edges and clusters in the graph, and minimizes the number of inter-cluster edges. Equation (4) ensures that an edge can be exclusively an intra- or inter-cluster edge. Equation (5) ensures that each vertex must belong to a cluster. Equation (6) ensures that an intra-cluster edge must have its both endpoints belonging to the same cluster. Equation (7) defines the variable C_j as a cluster head, if any vertex belongs to its respective cluster.

IV. Proposed Heuristic Algorithms

Algorithms 1 to 5 present five variations of Monte Carlo schemes that have been applied. The first variation assumes a fixed number of clusters ($\mathcal{C}$) and assigns each vertex to those clusters randomly. The second variation randomly picks a vertex which is assigned to its first neighbor cluster if exists, otherwise the vertex is made the head of a newly created cluster. The third variation randomly selects a vertex pair and if any cluster exists on either vertices' first neighbor, both vertices are assigned to that cluster. Otherwise both vertices are made head of two new clusters.

The fourth variation randomly picks an edge $e_i \in E$ and if any cluster exists on either end point's first neighbor, both end points are assigned to that cluster. Otherwise both endpoints are made head of two new clusters. This algorithm initializes $\sqrt{|NumberofVertex|}$ number of cluster and processes all edge-end-points by finding first neighbor cluster. In each of those variations the Kal (κ), CC (Ψ) and CPL (χ) is measured for those clusters.

Algorithm 1: MC-1

Input: Call Graph $G(V, E)$
Output: Clustering $\mathcal{C}$, Kal (κ), Clustering Coefficient (Ψ), Characteristics Path Length (χ)
Begin
Assume we have $n = \sqrt{|V|}$ number of clusters $C_1, C_2, C_3.....C_n$
for each vertex $v \in V$ **do**
  Generate a random number $i \in [0, |V|]$ for vertex v
  Assign v to cluster C_i
end for
Calculate κ, Ψ, χ using clustering scheme $\mathcal{C}$ and equations (1), (2) and (3) respectively
End

Algorithm 2: MC-2

Input: Call Graph $G(V, E)$
Output: Clustering $\mathcal{C}$, Kal(κ), Clustering Coefficient(Ψ), Characteristics Path Length (χ)
Begin
$\vartheta \leftarrow V$
while $\vartheta \neq \phi$ **do**
  Randomly pick vertex $v \in \vartheta$
  ω be the set of clusters, $\exists_{u\in\omega_j}(u, v) \in E$, $\forall\omega_j \in \omega$
  if $\omega = \phi$ **then**
    Create new cluster $\mathcal{C}_i$
    $\mathcal{C}_i \leftarrow \mathcal{C}_i \cup \{v\}$
  else
    Randomly pick $\mathcal{C}_i \in \omega$
    $\mathcal{C}_i \leftarrow \mathcal{C}_i \cup \{v\}$
  end if
  $\vartheta \leftarrow \vartheta \setminus \{v\}$
end while
Calculate κ, Ψ, χ using clustering scheme $\mathcal{C}$ and equations (1), (2) and (3) respectively
End

Algorithm 3: MC-3

Input: Call Graph $G(V, E)$
Output: Clustering $\mathcal{C}$, Kal(κ), Clustering Coefficient(Ψ), Characteristics Path Length (χ)
Begin
while All vertex $v \in V$ are not assigned to any cluster **do**
  Randomly pick vertex pair $\{v_1, v_2\} \in V$
  if $v_1 \in \mathcal{C}_i$ and v_2 unassigned to any cluster **then**
    $\mathcal{C}_i \leftarrow \mathcal{C}_i \cup \{v_2\}$
  else if v_1 unassigned to any cluster and $v_2 \in \mathcal{C}_j$ **then**
    $\mathcal{C}_j \leftarrow \mathcal{C}_j \cup \{v_1\}$
  else if v_1 and v_2 both unassigned to any clusters **then**
    for each $\nu_i \in \{v_1, v_2\}$ **do**
      ω be the set of clusters, $\exists_{u\in\omega_j}(u, \nu_i) \in E$, $\forall\omega_j \in \omega$
      if $\omega = \phi$ **then**
        Create new cluster $\mathcal{C}_i$
        $\mathcal{C}_i \leftarrow \mathcal{C}_i \cup \{\nu_i\}$
      else
        Randomly pick $\mathcal{C}_i \in \omega$
        $\mathcal{C}_i \leftarrow \mathcal{C}_i \cup \{\nu_i\}$
      end if
    end for
  end if
end while
Calculate κ, Ψ, χ using clustering scheme $\mathcal{C}$ and equations (1), (2) and (3) respectively
End

Algorithm 4: MC-4

Input: Call Graph $G(V, E)$
Output: Clustering $\mathcal{C}$, Kal(κ), Clustering Coefficient(Ψ), Characteristics Path Length (χ)
Begin
$\xi \leftarrow E$
for each edge $e \in \xi$ **do**
  Randomly pick an edge $e \in \xi$
  Assume v_1 and v_2 be the two end points of e
  if $v_1 \in \mathcal{C}_i$ and v_2 unassigned to any cluster **then**
    $\mathcal{C}_i \leftarrow \mathcal{C}_i \cup \{v_2\}$
  else if v_1 unassigned to any cluster and $v_2 \in \mathcal{C}_j$ **then**
    $\mathcal{C}_j \leftarrow \mathcal{C}_j \cup \{v_1\}$
  else if v_1 and v_2 both unassigned to any clusters **then**
    for each $\nu_i \in \{v_1, v_2\}$ **do**
      ω be the set of clusters, $\exists_{u \in \omega_j}(u, \nu_i) \in E$, $\forall \omega_j \in \omega$
      if $\omega = \phi$ **then**
        Create new cluster $\mathcal{C}_i$
        $\mathcal{C}_i \leftarrow \mathcal{C}_i \cup \{\nu_i\}$
      else
        Randomly pick $\mathcal{C}_i \in \omega$
        $\mathcal{C}_i \leftarrow \mathcal{C}_i \cup \{\nu_i\}$
      end if
    end for
  end if
  $\xi \leftarrow \xi \setminus \{e\}$
end for
Calculate κ, Ψ, χ using clustering scheme $\mathcal{C}$ and equations (1), (2) and (3) respectively
End

Algorithm 5: MC-5

Input: Call Graph $G(V, E)$
Output: Clustering $\mathcal{C}$, Kal(κ), Clustering Coefficient(Ψ), Characteristics Path Length (χ)
Begin
Fix initial number of clusters $\mathcal{C}_{i=1,\ldots,n}$ to $n = \sqrt{|V|}$
Randomly pick unique vertex $v_i \in V$, and a make one-to-one correspondence assignment of v_i to $\mathcal{C}_j$, where $i, j = 1, 2, 3 \ldots n$.
repeat
  $\xi_1 \leftarrow E$
  Randomly pick $\sqrt{|\xi_1|}$ edges from ξ_1 in ξ
  $\xi_1 \leftarrow \xi_1 \setminus \xi$
  repeat
    Randomly pick an edge $e \in \xi$
    Assume v_1 and v_2 be the two end points of e
    if $v_1 \in \mathcal{C}_i$ and v_2 unassigned to any cluster **then**
      $\mathcal{C}_i \leftarrow \mathcal{C}_i \cup \{v_2\}$
    else if v_1 unassigned to any cluster and $v_2 \in \mathcal{C}_j$ **then**
      $\mathcal{C}_j \leftarrow \mathcal{C}_j \cup \{v_1\}$
    else if v_1 and v_2 both unassigned to any clusters **then**
      for each $\nu_i \in \{v_1, v_2\}$ **do**
        ω be the set of clusters, $\exists_{u \in \omega_j}(u, \nu_i) \in E$, $\forall \omega_j \in \omega$
        if $\omega = \phi$ **then**
          Create new cluster $\mathcal{C}_i$
          $\mathcal{C}_i \leftarrow \mathcal{C}_i \cup \{\nu_i\}$
        else
          Randomly pick $\mathcal{C}_i \in \omega$
          $\mathcal{C}_i \leftarrow \mathcal{C}_i \cup \{\nu_i\}$
        end if
      end for
    end if
    remove the edge e from ξ
  until $\xi \neq \phi$
until all $v_i \in V$ are assigned to a $\mathcal{C}_j \in \mathcal{C}$
Calculate κ, Ψ, χ using clustering scheme $\mathcal{C}$ and equations (1), (2) and (3) respectively
End

Algorithm 6: Greedy-1

Input: Call Graph $G(V, E)$Characteristics Path Length
Output: Clustering $\mathcal{C}$, Kal(κ), Clustering Coefficient(Ψ), Characteristics Path Length (χ)
Begin
$\nu \leftarrow V$
for each vertex $v \in \nu$ in decreasing order of vertex degree **do**
  if $(v, u_i) \in E$ and $u_i \in \mathcal{C}_j$: for any $i = 1 \ldots |V|, j = 1 \ldots |\mathcal{C}|$ **then**
    $\mathcal{C}_j \leftarrow \mathcal{C}_j \cup \{v\}$
  else
    Create new cluster $\mathcal{C}_k$
    $\mathcal{C}_i \leftarrow \mathcal{C}_i \cup \{v\}$
  end if
  $\nu \leftarrow \nu \setminus \{v\}$
end for
Calculate κ, Ψ, χ using clustering scheme $\mathcal{C}$ and equations (1), (2) and (3) respectively
End

Algorithm 7: Greedy-2

Input: Call Graph $G(V, E)$
Output: Clustering $\mathcal{C}$, Kal(κ), Clustering Coefficient(Ψ), Characteristics Path Length (χ)
Begin
$\nu \leftarrow V$
for each vertex $v \in \nu$ in decreasing order of vertex degree **do**
  if $v \notin C_i$, $i = 1 \ldots |\mathcal{C}|$ **then**
    Set $\nu_1 := \phi$
    $\nu_1 \leftarrow \nu_1 \cup \{v\}$
    for each $(v, u_j) \in E$, $j = 1 \ldots |V|$ **do**
      if $u_j \notin C_k$, $k = 1 \ldots |\mathcal{C}|$ **then**
        $\nu_1 \leftarrow \nu_1 \cup \{u_j\}$
      end if
    end for
    Create new cluster $\mathcal{C}'$
    for each $(v_j) \in \nu_1$ **do**
      $\mathcal{C}' \leftarrow \mathcal{C}' \cup \{v_j\}$
    end for
    $\nu \leftarrow \nu \setminus \nu_1$
  end if
end for
Calculate κ, Ψ, χ using clustering scheme $\mathcal{C}$ and equations (1), (2) and (3) respectively
End

Algorithm 8: Greedy-3

Input: Call Graph $G(V, E)$
Output: Clustering $\mathcal{C}$, Kal(κ), Clustering Coefficient(Ψ), Characteristics Path Length (χ)
Begin
Fix initial number of clusters $\mathcal{C}_{i=1,\ldots,n}$ to $n = \sqrt{|V|}$
Pick unique vertex $v_i \in V$ in decreasing order of vertex degree and a make one-to-one correspondence assignment of v_i to $\mathcal{C}_j$, where $i, j = 1, 2, 3 \ldots n$.
for each edge $e \in E$ **do**
  Assume v_1 and v_2 be the two end points of e
  if $v_1 \in \mathcal{C}_i$ and v_2 unassigned to any cluster **then**
    $\mathcal{C}_i \leftarrow \mathcal{C}_i \cup \{v_2\}$
  else if v_1 unassigned to any cluster and $v_2 \in \mathcal{C}_j$ **then**
    $\mathcal{C}_j \leftarrow \mathcal{C}_j \cup \{v_1\}$
  end if
end for
Calculate κ, Ψ, χ using clustering scheme $\mathcal{C}$ and equations (1), (2) and (3) respectively
End

Algorithm 6 to 8 present greedy based huristics that have been applied. The first algorithm selects each of vertices $v_i \in V$ in descending order of their nodal degree. After selection, a vertex is assigned to its adjacent neighboring cluster if exists. Otherwise that vertex is made the head of a new cluster.

TABLE II. NUMBER OF POTENTIAL CLASSES GENERATED BY DIFFERENT HURISTICS ALGORITHMS

Data	MC1	MC2	MC3	MC4
BFT	3	2	3	1
RBIo	7	7	7	2
Synth	2	2	2	2
Data	MC5	G1	G2	G3
BFT	3	5	8	7
RBIo	8	5	14	8
Synth	3	1	2	3

The second algorithm picks each of vertices $v_i \in V$ in descending order based on nodal degree and allocate the adjacent vertex including itself to a new cluster. Finally the third algorithm initializes $\sqrt{|NumberofVertex|}$ number of cluster and select each of vertices $v_i \in V$ in descending order based on nodal degree. In this algorithm, if one edge-end-point is exists at any cluster than assign the other end-points to the same cluster.

V. EXPERIMENTAL SETUP AND RESULTS

Algorithms presented in Section IV were implemented using C++ programming language on 32bit Linux Mint15 machine with Intel Core-i3 processor, 4GB RAM. We have reported experimental results on 3 problem instances. Instance *BFT* and *RBIo* [4] have been generated from two scientific research softwares, and the other instance *Synthetic1* is synthetically generated. Number of functions and number of calls in the dataset *BFT*, *RBIo* and *Synthetic1* are (14, 31), (61, 372) and (6, 7) respectively. Table II reports the number of potential classes generated by different huristics algorithms for those experimental datasets.

Table III reports the computational running time in microsecond. MC algorithms were executed 1000 times on each of the instances and the total execution times have been presented. According to instance *BFT* MC-3 is 34.66% faster in average than other MC algorithms. Greedy algorithms, since they has been executed only once, convincingly outperformed MC algorithms. In terms of computational time Greedy-2 is 0.59%, and 14.20% faster than Greedy-1 and Greedy-3 repectively.

The results of all MC algorithms for instance *BFT* is presented in Fig. 1; subsection (a),(b),(c) represent CC, CPL and Kal index respectively. Output results are displayed using boxplot diagram where every box focuses the majority portion of the proposed result. The solution value does not vary significantly among all the instance run except MC-1 algorithm. For instance *BFT* the average and best result of (Ψ), (χ) and (κ) are (0.22194, 0.458333), (1.642, 0.7), and (5, 17) respectively. The largest and shortest running times of this instance are 19819 microseconds and 11855 microseconds.

Fig. 2 represents the best and average result of MC algorithms in contrast to Greedy algorithms. We see that the Greedy algorithms produce better result than the average performances of MC algorithms, and are almost closest of MC algorithms best results. Sign $(*)$, (x), and $(+)$ denote the average result of MC, best result of MC and optimal result of Greedy algorithms respectively.

The comparison results between MC and Greedy algorithms on the experimental instances are given in Table IV, V and VI which are prepared from instance *BFT*, instance *RBIo*, and instance *Synthetic* respectively. Each table presents the values of (Ψ), (χ) and (κ) of every huristics algorithms. Those tables and diagram show that the proposed Greedy algorithms are 12.8%, 31.02%, and 25.73% better than proposed Monte Carlo algorithms in terms of CC (Ψ), CPL (χ) and Kal index (κ) respectively.

TABLE III. RUNNING TIME (IN MICROSECONDS)

Data	MC1	MC2	MC3	MC4
BFT	15152	12316	11855	19819
RBIo	1304418	1510649	1373883	3681185
Synth	12747	8248	29793	33737
Data	MC5	G1	G2	G3
BFT	16571	193	169	170
RBIo	2146621	42238	23232	16598
Synth	8822	56	50	48

TABLE IV. MC1-MC5 AND GREEDY1-GREEDY3 ALGORITHMS ON BFT

Algorithm	CC		CPL		KL	
	Average	Best	Average	Best	Average	Best
MC1	0.23542	0.23148	3.0810	2.0873	-5	5
MC2	0.23738	0.13888	1.224	1.7651	5	14
MC3	0.22513	0.16666	1.182	3.1868	5	14
MC4	0.19129	0.23809	1.613	1.026	11	16
MC5	0.2219	0.11111	1.197	1.3481	5	13
Greedy1	0.324074		0.98666		13	
Greedy2	0.3925		0.76667		1	
Greedy3	0.38888		1.01334		7	

VI. CONCLUSION AND FUTURE WORK

The work presented in this paper has the following areas of development: Firstly, modeling the SP code as a call graph does not consider the global variables and function parameters (as has been used by [13]), which could be useful intuitions for the design of classes. Secondly, approaches presented in this paper do not address discovery of polymorphism and inheritance of classes. Thirdly, we have not considered how many times one function call the other.

Further research on the problem may be directed towards

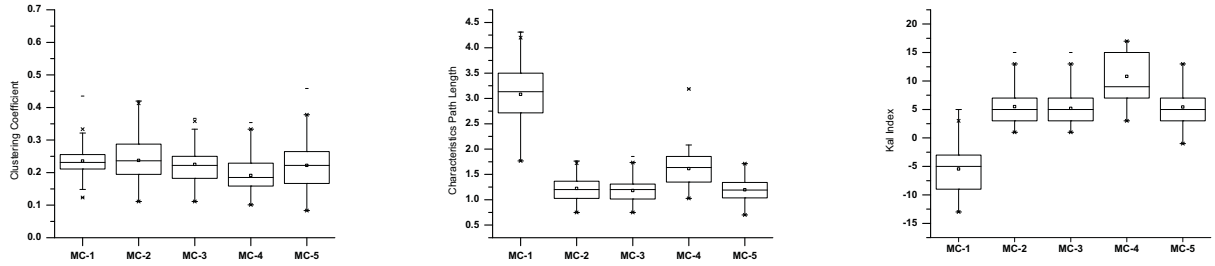
TABLE V. MC1-MC5 AND GREEDY1-GREEDY3 ALGORITHMS RBIO

Algorithm	CC		CPL		KL	
	Average	Best	Average	Best	Average	Best
MC1	0.3321	0.31129	3.0704	2.93199	-98	-81
MC2	0.3189	0.24926	1.0818	1.17601	3	71
MC3	0.3393	0.13958	1.0481	1.22531	-3	73
MC4	0.2114	0.17372	1.8273	2.09649	78	101
MC5	0.2954	0.26033	1.2672	1.25219	25	61
Greedy1	0.280905		1.34956		105	
Greedy2	0.219878		0.736625		49	
Greedy3	0.302799		1.41524		11	

TABLE VI. MC1-MC5 AND GREEDY1-GREEDY3 ALGORITHMS ON SYNTHETIC1

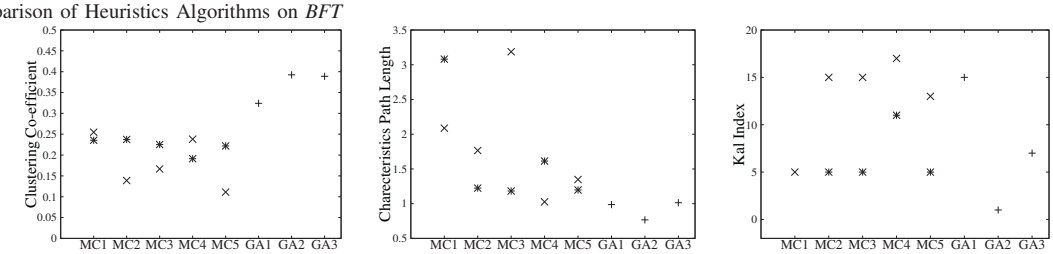
Algorithm	CC		CPL		KL	
	Average	Best	Average	Best	Average	Best
MC1	0.7693	0.388889	1.6265	0.9	0	7
MC2	0.81077	0.777778	1.1078	1.8	4	7
MC3	0.80555	0.777778	1.1408	1.8	4	7
MC4	0.7844	0.777778	1.476	1.8	6	7
MC5	0.8145	0.777778	0.932	1	3	5
Greedy1	0.777		1.8		7	
Greedy2	0.83333		1.16666		3	
Greedy3	0.6667		0.7037		1	

Fig. 1. Monte Carlo Algorithms on *BFT*



(a) Clustering Coefficient(Ψ) (b) Characteristics Path Length(χ) (c) KAL index (κ)

Fig. 2. Comparison of Heuristics Algorithms on *BFT*



(a) Clustering Coefficient(Ψ) (b) Charecteristics Path Length(χ) (c) KAL index (κ)

the using numerical optimization problem solver software to run the model presented in Section III to get benchmarks results; testing the proposed larger legacy code instances, and empirically validating the code by professional OOP experts. We intend to use meta-heuristic solution space search algorithms on the problem in the next step to get better optimized results.

ACKNOWLEDGMENT

This research has been supported by The University Grant Commission, Bangladesh under the Dhaka University Teachers Research Grant No-Regi/Admn-3/2012-2013/13190.

Thanks to Mr. Md. Selim of IITDU Optimization Research Group for his total cooperation. Our sincere gratitude to Dr. Shahadat Hossain, Associate Professor, Dept. Math & Computer Science, Univ. of Lethbridge, AB, Canada for presenting this problem to us. Thanks to Mr. Tahsin Zulkarnine, Web Application Developer, Commerx Corporation, Calgary, AB, Canada for providing experimental data.

REFERENCES

[1] K. Chisolm and J. Lisonbee, "The use of computer language compilers in legacy code migration," in *IEEE Systems Readiness Technology Conference (AUTOTESTCON'99)*. IEEE, 1999, pp. 137–145.

[2] B. G. Ryder, "Constructing the call graph of a program," *Software Engineering, IEEE Transactions on*, no. 3, pp. 216–226, 1979.

[3] Y. Terashima and K. Gondow, "Static call graph generator for C++ using debugging information," in *14th Asia-Pacific Software Engineering Conference (APSEC 2007)*. IEEE, 2007, pp. 127–134.

[4] S. Hossain and A. T. Zulkarnine, "Design structure of scientific software–a case study," in *13th International DSM Conference*. MIT, Cambridge, Massachusetts, USA: IEEE, 2011, pp. 129–141.

[5] S. R. Chidamber and C. F. Kemerer, "A metrics suite for object oriented design," *IEEE Transactions on Software Engineering*, vol. 20, no. 6, pp. 476–493, 1994.

[6] P. O. Asagba and E. Ogheneovo, "A comparative analysis of structured and object-oriented programming methods," *Journal of Applied Sciences and Environmental Management*, vol. 11, no. 4, pp. 42–46, 2007.

[7] L. A. Wolsey, "Integer programming," *IIE Transactions*, vol. 32, no. 273-285, pp. 2–58, 2000.

[8] D. J. Watts and S. H. Strogatz, "Collective dynamics of small-worldnetworks," *nature*, vol. 393, no. 6684, pp. 440–442, 1998.

[9] D. Braha and Y. Bar-Yam, "The statistical mechanics of complex product development: empirical and analytical results," *Management Science*, vol. 53, no. 7, pp. 1127–1145, 2007.

[10] S. E. Schaeffer, "Graph clustering," *Computer Science Review*, vol. 1, no. 1, pp. 27–64, 2007.

[11] H. M. Sneed and E. Nyary, "Extracting object-oriented specification from procedurally oriented programs," in *2nd Working Conference on Reverse Engineering*. Toronto, Ont., Canada: IEEE, 1995, pp. 217–226.

[12] O. Maqbool and H. A. Babri, "Hierarchical clustering for software architecture recovery," *IEEE Transactions on Software Engineering*, vol. 33, no. 11, pp. 759–780, 2007.

[13] V. Dineshkumar and J. Deepika, "Code to design migration from structured to object oriented paradigm," *International Journal of Information and Communication Technology Research*, vol. 1, 2011.

[14] Y. Zhou, H. Cheng, and J. X. Yu, "Graph clustering based on structural/attribute similarities," *VLDB Endowment*, vol. 2, no. 1, pp. 718–729, 2009.

[15] C. C. Aggarwal, J. L. Wolf, P. S. Yu, C. Procopiuc, and J. S. Park, "Fast algorithms for projected clustering," in *ACM SIGMOD international conference on Management of data*, ser. SIGMOD '99. New York, NY, USA: ACM, 1999, pp. 61–72.

[16] K.-L. Wu and M.-S. Yang, "Alternative c-means clustering algorithms," *Pattern recognition*, vol. 35, no. 10, pp. 2267–2278, 2002.

[17] J. C. Bezdek and N. R. Pal, "Some new indexes of cluster validity," *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, vol. 28, no. 3, pp. 301–315, 1998.

[18] J. C. Dunn, "A fuzzy relative of the isodata process and its use in detection compact well-separated clusters," *Journal of Cybernetics*, vol. 3, pp. 32–57, 1973.

Joint Routing, Scheduling, and Physical Network Coding in Fixed Wireless Multihop Networks

Md. Forkan Uddin, Fateha Noor, and Nazia Mehnaz
Dept. of Electrical and Electronic Engineering
Bangladesh University of Engineering and Technology
Dhaka, Bangladesh
Email: mforkanuddin@eee.buet.ac.bd, fatiha_noor@yahoo.com, araninazia@yahoo.com

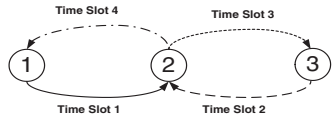
***Abstract*—We study joint routing, scheduling, and physical network coding (PNC) in fixed wireless multihop networks under a *physical interference model*. The objective is to quantify the throughput gain provided by PNC in wireless mesh networks. We formulate a cross-layer optimization framework to determine the optimal max-min throughput of the flows and the optimal configuration of the routing and scheduling parameters in multi-hop wireless networks with PNC. We solve the problem for several small to medium size wireless mesh networks. Our numerical results show that PNC provides a significant throughput gain in wireless mesh networks at low transmission power. This throughput gain is significantly higher than the throughput gain of the XOR-like network coding.**

***Index Terms*—Cross-layer optimization, throughput, routing, scheduling, physical network coding, link layer network coding.**

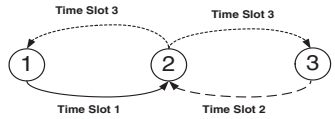
I. Introduction

Network coding (NC) has emerged as a promising technique in communication systems which manipulates data inside the network to increase throughput, reduce delay, and improve robustness [1], [2]. It has been used in both wireless and wireline networks in many context to improve network performance. In wireless networks, NC has been used for (i) end-to-end multicasting [3], (ii) end-to-end unicasting [4], (iii) at the link layer [5], [6], and (iv) physical layer transmission [7], [8]. However, link layer NC [6] and physical network coding (PNC) [8] have attracted a lot of attention from the wireless research community since these NC techniques are simpler than the end-to-end NC techniques.

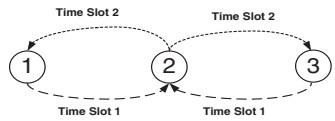
Traditionally, a node cannot transmit and receive multiple packets at the same time. In Fig. 1(a), for example, assume that node 1 (resp. node 3) needs to send a packet of flow f_1 (resp. f_2) to node 3 (resp. node 1) through the intermediate node 2. Without any NC, four time slots are necessary to complete the transmissions since a node usually cannot transmit and receive at the same time (See Fig. 1(a)). On the other hand, if link layer NC, i.e., XOR-like NC [6] is applied as shown in Fig. 1(b), after the first two time slots, one packet from each flow will be available at node 2, and then it will transmit both packets simultaneously by XORing them in the third time slot. Node 1 (resp. node 3) can then decode the packet intended for itself by XORing the packet it sent together with the received XORed packet. Thus, three time slots are necessary for system with XOR-like NC. When PNC [8] is applied, as shown in Fig. 1(c), in the first slot, both nodes 1 and 3 transmit their packets simultaneously and hence, node 2 cannot decode the information bits separately. However, node 2 can store the information which it received in the first slot and then it can transmit the information in the second time slot. Nodes 1 and 3 can then decode the packet intended for itself. Thus, only two time slots are required to complete the transmissions and hence, it is expected that PNC will improve throughput significantly in multi-hop wireless networks.



(a) Traditional



(b) XOR-like network coding



(c) Physical network coding

Fig. 1. Different types of transmission schemes.

Wireless mesh networks are considered as promising multi-hop wireless networks for future Internet services due to lower costs of infrastructure, easier administration, better coverage, and lower power consumption. Throughput is one of the important performance metrics in wireless mesh networks which significantly depends on the joint configuration of routing and medium access control (MAC) parameters [9], [10]. However, how to jointly configure the network parameters when PNC is enabled is unknown. In [11], the authors analytically show that 100% gain can be achieved by using PNC in a linear network with balanced traffic. However, there is no clear indication so far on how much throughput can be improved in wireless mesh network by enabling PNC.

In this paper, we study joint routing, scheduling, and physical network coding (JRS-PNC) problem in wireless networks to maximize the minimum throughput of the flows. Our objective is to provide insight on the throughput gains obtained

by PNC in wireless mesh networks. We consider a *physical interference model* based on signal to interference plus noise ratio (SINR) since it is a more realistic interference model for wireless interference [12]. The contributions in this paper are as follows:

- We formulate a JRS-PNC optimization problem to determine the weighted max-min throughputs of the flows and the optimal configuration of routing and scheduling parameters in wireless networks with PNC.
- The JRS-PNC optimization problem is very large scale NP-hard Linear Program (LP). We solve the JRS-PNC optimization problem numerically for several small mesh network scenarios due to the complexity of the problem. We also solve the joint problem by limiting PNC to only bidirectional flows which has lower complexity. We find that almost similar gains can be achieved by limiting PNC to only bidirectional flows. We then solve the optimization problem with bidirectional PNC for several medium size wireless mesh networks and find that PNC provides a significant throughput gain at a low transmit power. We also compare the throughput performance of XOR-like NC and PNC. We find that PNC provides a significantly higher gain than the XOR-like NC.

The rest of the paper is organized as follows. Related work is reviewed in Section II. We formulate the optimization problem in Section III. In Section IV, we study the cross-layer optimization problem for wireless mesh networks. Section V concludes the paper.

II. Related Work

Network coding was first introduced in [1] for multicast applications in wireline networks [1]. Since then a large body of work on NC has been done for end-to- end multicast and unicast applications in wireline and wireless networks [3], [4]. End-to-end NC is very complex and difficult to implement. On the other hand, the link layer NC and PNC are simpler than the end-to-end NC. The authors in [5] introduced link layer NC, i.e., XOR-type NC for unicast applications. Since then a lot of research have been done on link layer NC [6], [13], [14]. The concept of PNC originally proposed in [8] based on mapping and demodulation. Since then the same authors in [11] study the throughput performance of PNC in linear network with balanced traffic and demonstrate that the throughput gain obtained by PNC is 100%.

In [15], Jain *et al.* formulate the original optimization framework for the joint routing and scheduling (JRS) problem for any given wireless network with a traditional transmission scheme. In [13], [14], the authors study the cross-layer optimization problem with XOR-like link layer NC. In [13], the authors study joint routing, scheduling, and XOR-like NC (JRS-XNC) under a simplistic interference model. In [14], the authors study the JRS-XNC problem under the *physical interference model*. So far, cross-layer optimization problem in multi-hop wireless networks with XOR-like NC is well studied. However, the study on cross-layer optimization of multi-hop wireless networks with PNC is not yet presented.

III. Problem Formulations

A. System Model

1) Network Topology and Flows: We consider a wireless network consisting of N stationary nodes with known locations. The set of nodes is denoted by $\mathcal{N}$. The transmit power of the nodes is equal and fixed. The transmit power is denoted by P_t. There are F data flows in the network, belonging to set $\mathcal{F}$. A data flow f is characterized by its source f^s, its destination f^d and its rate λ_f. The rates of flows satisfy:

$$\lambda_f = w_f \lambda \quad \forall f \in \mathcal{F}, \tag{1}$$

where λ is a common base throughput and w_f is a known weight. Our objective is to maximize the common base throughput λ by maintaining a preset traffic rate ratio for the flows. We assume that each node uses omni- directional antenna and cannot transmit and receive simultaneously.

2) Physical network coding Model: We assume that all the nodes are capable to perform PNC. We consider a PNC model as illustrate in Fig. 1(c). Thus, PNC opportunity at a node involves of two packets and these packets must enter through a pair of incoming links and leave through an opposite pair of outgoing links.

3) Channel and Interference Models: The channel gain between a transmitter and a receiver is assumed to be time invariant. The channel gain between nodes n_1 and n_2, $G_{n_1 n_2}$, is given by $(d_{n_1 n_2}/d_0)^{-\eta}$, where $d_{n_1 n_2}$ is the distance between nodes n_1 and n_2, d_0 is a reference distance in the far field of the transmit antenna, and η is the path loss exponent. We assume that all the nodes use the same modulation and coding scheme and the minimum required SINR for a successful decoding of a packet is γ. The normalized transmission rate for the modulation and coding scheme is assumed to be one. We represent a directed link l as (l^o, l^d), where l^o and l^d are the originating and destination nodes of the link. A directed link between nodes n_1 and n_2 exists if they can communicate in the absence of interference, i.e., if the signal to noise ratio (SNR) for the link is not less than γ, i.e.,

$$\frac{G_{n_1 n_2} P_t}{N_0} \geq \gamma \tag{2}$$

where N_0 is the background noise power at a receiver. Let the number of directed links in the network is L and their set is $\mathcal{L}$. Denote the sets of links coming into and going out of node n by $\mathcal{L}_n^I$ and $\mathcal{L}_n^O$. A transmission on link l without PNC is considered as successful if the received SINR at l^d is not less than γ for the whole packet duration, i.e.,

$$\frac{G_{l^o l^d} P_t}{N_0 + \sum_{l' \neq l} G_{l'^o l^d} P_t Y_{l'}} \geq \gamma \tag{3}$$

where $Y_{l'}$ is a binary variable being equal to 1 if there is a transmission on link l' at the same time and 0 otherwise. In [8], the authors show that the SINR requirement for a transmission with PNC is approximately equal to the SINR requirement of a transmission without PNC for a given bit error rate. Thus, transmissions on links l_i and l_j, $l_i^d = l_j^d$, with PNC are considered as successful if

$$\frac{G_{l_i^o l_i^d} P_t}{N_0 + \sum_{l' \neq l_i, l_j} G_{l'^o l_i^d} P_t Y_{l'}} \geq \gamma \tag{4}$$

and

$$\frac{G_{l_j^o l_j^d} P_t}{N_0 + \sum_{l' \neq l_i, l_j} G_{l'^o l_j^d} P_t Y_{l'}} \geq \gamma. \tag{5}$$

Similarly, transmissions on links $\bar{l}_i$ and $\bar{l}_j$, $\bar{l}_i^o = \bar{l}_j^o$, with PNC are considered as successful if

$$\frac{G_{\bar{l}_i^o \bar{l}_i^d} P_t}{N_0 + \sum_{l' \neq \bar{l}_i, \bar{l}_j} G_{l'^o \bar{l}_i^d} P_t Y_{l'}} \geq \gamma \tag{6}$$

and

$$\frac{G_{\bar{l}_j^o \bar{l}_j^d} P_t}{N_0 + \sum_{l' \neq \bar{l}_i, \bar{l}_j} G_{l'^o \bar{l}_j^d} P_t Y_{l'}} \geq \gamma \tag{7}$$

where $\bar{l}$ is the opposite link of link l, i.e., $l^o = \bar{l}^d$ and $l^d = \bar{l}^o$. Note that there should be a condition that, at a node, the received traffic rate with PNC is equal to the transmit data rate with PNC. This issue is addressed later in Section III-B during problem formulation.

4) Scheduling: A set of links that can be scheduled without any collision is defined as an independent set (ISET) in the literature [9], [15]. We define that a set of links v is an ISET if

$$\frac{P_t G_{l^o l^d}}{\sum_{l' \in v \setminus l, l'^d \neq l^d, l'^o \neq l^o} P_t G_{l'^o l^d} + N_0} \geq \gamma \quad \forall l \in v; \tag{8}$$

$$\sum_{l' \in v \setminus l} 1_{\{l'^d = l^d\}} + \sum_{l' \in v \setminus l} 1_{\{l'^o = l^o\}} \leq 1 \quad \forall l \in v. \tag{9}$$

If an ISET v consists of two links with same destination, i.e., two links l_i and l_j belong to the set v and $l_i^d = l_j^d$ then the transmitters of those two links will transmit packets simultaneously with PNC during the schedule of the ISET. Further, if the ISET v consists of two links with same origin then the origin of the two links will transmit the stored information with PNC which is received earlier. The transmitters of the rest of links of the ISET v, i.e., those do not have the same destination or origin, will transmit packets without PNC.

B. Joint Routing, Scheduling and Physical Network Coding

Let the set of all the ISETs in the network is $\mathcal{I}$. Denote by $\mathcal{I}_{l_i}^{wnc}$ the set of ISETs that includes link l_i and the link l_i to be scheduled to transmit without network coding, i.e., $l_i^o \neq l_j^o$ and $l_i^d \neq l_j^d$ $\forall l_j \in v \setminus l_i$, $\forall v \in \mathcal{I}_{l_i}^{wnc}$. Denote by $\mathcal{I}_{l_i,l_j}^{pnc1}$ the set of ISETs that includes links l_i and l_j, $l_i^d = l_j^d$ and the links l_i and l_j to be scheduled to transmit with PNC. Similarly, we denote the set of ISETs that includes links l_i and l_j, $l_i^o = l_j^o$ by $\mathcal{I}_{l_i,l_j}^{pnc2}$ where the links l_i and l_j to be scheduled to transmit with PNC. Let α_v be the fraction of time that the ISET v is scheduled such that $\sum_{v \in \mathcal{I}} \alpha_v = 1$ and $\boldsymbol{\alpha}$ be the vector of the α_v variables. Let the rate of flow f_i on link l_i without NC is c_{f_i,l_i}^{wnc}. Thus,

$$\sum_{f_i \in \mathcal{F}} c_{f_i,l_i}^{wnc} = \sum_{v \in \mathcal{I}_{l_i}^{wnc}} \alpha_v \quad \forall n \in \mathcal{N}, \forall l_i \in \mathcal{L}_n^O. \tag{10}$$

We denote the matrix of the c_{f_i,l_i}^{wnc} variables by $\mathbf{c}^{wnc}$. Further, let the rate of flows f_i and f_j on links l_i and l_j, $l_i^d = l_j^d$, is $c_{f_i,l_i,f_j,l_j}^{pnc1}$ and the rate of flows f_i and f_j on links l_i and l_j, $l_i^o = l_j^o$, is $c_{f_i,l_i,f_j,l_j}^{pnc2}$ for transmissions with PNC. Thus,

$$\sum_{f_i \in \mathcal{F}, f_j \in \mathcal{F}, f_i \neq f_j} c_{f_i,l_i,f_j,l_j}^{pnc1} = \sum_{v \in \mathcal{I}_{l_i,l_j}^{pnc1}} \alpha_v$$
$$\forall n \in \mathcal{N}, \forall l_i \in \mathcal{L}_n^I, \forall l_j \in \mathcal{L}_n^I, l_i < l_j. \tag{11}$$

and

$$\sum_{f_i \in \mathcal{F}, f_j \in \mathcal{F}, f_i \neq f_j} c_{f_i,l_i,f_j,l_j}^{pnc2} = \sum_{v \in \mathcal{I}_{l_i,l_j}^{pnc2}} \alpha_v$$
$$\forall n \in \mathcal{N}, \forall l_i \in \mathcal{L}_n^O, \forall l_j \in \mathcal{L}_n^O, l_i < l_j. \tag{12}$$

Denote the matrices of the $c_{f_i,l_i,f_j,l_j}^{pnc1}$ and $c_{f_i,l_i,f_j,l_j}^{pnc2}$ variables by $\mathbf{c}^{pnc1}$ and $\mathbf{c}^{pnc2}$, respectively.

Combining the transmissions with and without PNC, the effective rate of flow f_i on link l_i $\forall fi \in \mathcal{F}, \forall n \in \mathcal{N}, \forall l_i \in \mathcal{L}_n^O$ can be written as

$$\begin{aligned} c_{f_i,l_i} &= c_{f_i,l_i}^{wnc} + \sum_{f_j \in \mathcal{F}, f_j \neq f_i, l_j \in \mathcal{L}_n^I, l_i < l_j} c_{f_i,l_i,f_j,l_j}^{pnc1} \\ &+ \sum_{f_j \in \mathcal{F}, f_j \neq f_i, l_j \in \mathcal{L}_n^I, l_i > l_j} c_{f_j,l_j,f_i,l_i}^{pnc1} + \sum_{f_j \in \mathcal{F}, f_j \neq f_i, l_j \in \mathcal{L}_n^O, l_i < l_j} c_{f_i,l_i,f_j,l_j}^{pnc2} \\ &+ \sum_{f_j \in \mathcal{F}, f_j \neq f_i, l_j \in \mathcal{L}_n^O, l_i > l_j} c_{f_j,l_j,f_i,l_i}^{pnc2}. \end{aligned} \tag{13}$$

Note that to perform PNC among the traffic of two flows and to transmit on the two outgoing links of a node, the traffic of the two flows must be received by the node through the opposite links of those two links with PNC. This is what we called the PNC constraints and they can be written as

$$c_{f_i,l_i,f_j,l_j}^{pnc1} = c_{f_j,\bar{l}_i,f_i,\bar{l}_j}^{pnc2} \quad \forall n \in \mathcal{N}, \forall l_i \in \mathcal{L}_n^I, \forall l_j \in \mathcal{L}_n^I, l_i < l_j. \tag{14}$$

Let the traffic rate of flow f_i on link l_j is y_{f_i,l_i}. Denote the matrix of the y_{f_i,l_i} variables by $\mathbf{y}$. To maximize the minimum throughput of the flows, the JRS-PNC problem can be formulated as in (15)-(24).

$$\max_{\mathbf{c}^{wnc}, \mathbf{c}^{pnc1}, \mathbf{c}^{pnc2}, \boldsymbol{\alpha}, \mathbf{y},} \quad \lambda \tag{15}$$

$$\text{s.t.} \sum_{l_i \in \mathcal{L}_n^O} y_{f_i,l_i} - \sum_{l_i \in \mathcal{L}_n^I} y_{f_i,l_i} = \begin{cases} w_{f_i}\lambda & \text{if } n = f_i^s \\ -w_{f_i}\lambda & \text{if } n = f_i^d \\ 0 & \text{otherwise} \end{cases}$$
$$\forall n \in \mathcal{N}, \forall f_i \in \mathcal{F} \tag{16}$$

$$\begin{aligned} y_{f_i,l_i} &\leq c_{f_i,l_i}^{wnc} + \sum_{f_j \in \mathcal{F}, f_j \neq f_i, l_j \in \mathcal{L}_n^I, l_i < l_j} c_{f_i,l_i,f_j,l_j}^{pnc1} \\ &+ \sum_{f_j \in \mathcal{F}, f_j \neq f_i, l_j \in \mathcal{L}_n^I, l_i > l_j} c_{f_j,l_j,f_i,l_i}^{pnc1} \\ &+ \sum_{f_j \in \mathcal{F}, f_j \neq f_i, l_j \in \mathcal{L}_n^O, l_i < l_j} c_{f_i,l_i,f_j,l_j}^{pnc2} \end{aligned}$$

$$+ \sum_{f_j \in \mathcal{F}, f_j \neq f_i, l_j \in \mathcal{L}_n^O, l_i > l_j} c^{pnc2}_{f_j,l_j,f_i,l_i} \quad \forall f_i \in \mathcal{F}, \forall l_i \in \mathcal{L} \quad (17)$$

$$\sum_{f_i \in \mathcal{F}} c^{wnc}_{f_i,l_i} = \sum_{v \in \mathcal{I}^{wnc}_{l_i}} \alpha_v \quad \forall n \in \mathcal{N}, \forall l_i \in \mathcal{L}_n^O \quad (18)$$

$$\sum_{f_i \in \mathcal{F}, f_j \in \mathcal{F}, f_i \neq f_j} c^{pnc1}_{f_i,l_i,f_j,l_j} = \sum_{v \in \mathcal{I}^{pnc1}_{l_i,l_j}} \alpha_v$$
$$\forall n \in \mathcal{N}, \forall l_i \in \mathcal{L}_n^I, \forall l_j \in \mathcal{L}_n^I, l_i < l_j \quad (19)$$

$$\sum_{f_i \in \mathcal{F}, f_j \in \mathcal{F}, f_i \neq f_j} c^{pnc2}_{f_i,l_i,f_j,l_j} = \sum_{v \in \mathcal{I}^{pnc2}_{l_i,l_j}} \alpha_v$$
$$\forall n \in \mathcal{N}, \forall l_i \in \mathcal{L}_n^O, \forall l_j \in \mathcal{L}_n^O, l_i < l_j \quad (20)$$

$$c^{pnc1}_{f_i,l_i,f_j,l_j} = c^{pnc2}_{f_j,\bar{l}_i,f_i,\bar{l}_j}$$
$$\forall n \in \mathcal{N}, \forall l_i \in \mathcal{L}_n^I, \forall l_j \in \mathcal{L}_n^I, l_i < l_j \quad (21)$$

$$\sum_{v \in \mathcal{I}} \alpha_v \leq 1 \quad (22)$$

$$0 \leq \lambda, \mathbf{y}, \mathbf{c}^{wnc}, \mathbf{c}^{pnc1}, \mathbf{c}^{pnc2} \quad (23)$$

$$0 \leq \boldsymbol{\alpha} \leq 1. \quad (24)$$

The objective function in (15) is to ensure that λ is maximized. The flow conservation constraints in (16) guarantee that the outgoing and incoming traffic of a flow are equal at each intermediate node, that the outgoing traffic of a flow is equal to the source rate at the source node, and that the incoming traffic of a flow is equal to the source rate at the destination node. The link rate constraints in (17) ensure that the traffic rate on a link is not larger than the link rate. The constraints in (18)- (20) relate the link rates with the scheduling of different ISETs. The constraints in (21) ensure that a node performs transmissions/receptions with PNC as much as possible. The equality constraint in (22) relates the scheduling of all ISETs and ensure that the total scheduling time is not more than one. The constraints in (23) and (24) define the range of the variables.

The JRS-PNC problem in (15)-(24) is a LP. However, it is a very large scale NP-hard problem. The number of variables and the number of constraints are in the order of $O(F^2L^2)$. Although the number of flows linearly grows with the size of a network, the number of links grows exponentially with the size of the network. The maximum number of directed links for a network of size N is $N(N-1)$. The number of ISETs also grows exponentially with the size of links. The number of ISETs depends on the number of links and the maximum spatial reuse of scheduling in the network. The maximum number of ISETs is in the order of $O(L^Q)$ where Q is the maximum spatial reuse of scheduling.

C. Simplification of JRS-PNC Problem

Define two flows f and g are bidirectional if $f^s = g^d$ and $f^d = g^s$. One can restrict PNC to bidirectional flows to reduce the complexity of the JRS-PNC optimization problem in (15)-(24). Denote the bidirectional flow of flow $f_i \in \mathcal{F}$ by $\overline{f}_i$. Then one can formulate the joint routing, scheduling, (bi-directional) physical network coding (JRS-BiPNC) optimization problem, by replacing $f_j \in \mathcal{F}$ with $\overline{f}_i$ in all the constraints in the JRS-PNC problem. Note that, by using PNC on bidirectional flows, the number of constraints and variables of the optimization problem reduce to $O(FL^2)$ from $O(F^2L^2)$.

D. Joint Routing and Scheduling without PNC

In [15], Jain *et al.* formulate the original optimization framework for the JRS problem. For a system without PNC, a set of links v is an ISET if

$$\frac{P_t G_{l^o l^d}}{\sum_{l' \in v \setminus l} P_t G_{l'^o l^d} + N_0} \geq \gamma \quad \forall l \in v. \quad (25)$$

Enumerating the ISETs using the above conditions, one can formulate the JRS problem from the JRS-PNC problem in (15)- (24) with the following steps:

1) Replace the link rate constraints in (17) with
$$y_{f_i,l_i} \leq c^{wnc}_{f_i,l_i} \quad \forall f_i \in \mathcal{F}, \forall l_i \in \mathcal{L}. \quad (26)$$
2) Remove the constraints in (19)-(21).
3) Replace the constraints in (23) with
$$0 \leq \lambda, \mathbf{y}, \mathbf{c}^{wnc}. \quad (27)$$

E. Joint Routing, Scheduling, and Bidirectional XOR-like Network Coding

For a system with XOR-like NC, a set of links v is an ISET if [14]

$$\frac{P_t G_{l^o l^d}}{\sum_{l' \in v \setminus l, l'^o \neq l^o} P_t G_{l'^o l^d} + N_0} \geq \gamma \quad \forall l \in v; \quad (28)$$

$$\sum_{l' \in v \setminus l} 1_{\{l'^o = l^o\}} \leq 1 \quad \forall l \in v. \quad (29)$$

If the ISET v consists of two links with same origin then the origin of the two links will transmit packets with XOR NC. The transmitters of the rest of links of the ISET v will transmit packets without any NC. Enumerating the ISETs using the conditions in (28) and (29), one can formulate the joint routing, scheduling, (bidirectional) XOR network coding (JRS-BiXNC) problem from the JRS-PNC problem in (15)-(24) with the following steps:

1) Replace the link rate constraints in (17) with
$$y_{f_i,l_i} \leq c^{wnc}_{f_i,l_i} + \sum_{f_j \in \mathcal{F}, f_j \neq f_i, l_j \in \mathcal{L}_n^O, l_i < l_j} c^{pnc2}_{f_i,l_i,f_j,l_j} + \sum_{f_j \in \mathcal{F}, f_j \neq f_i, l_j \in \mathcal{L}_n^O, l_i > l_j} c^{pnc2}_{f_j,l_j,f_i,l_i} \quad \forall f_i \in \mathcal{F}, \forall l_i \in \mathcal{L}. \quad (30)$$
2) Remove the constraints in (19) and (21).
3) Replace the constraints in (23) with
$$0 \leq \lambda, \mathbf{y}, \mathbf{c}^{wnc}, \mathbf{c}^{pnc2}. \quad (31)$$
4) Include the following NC constraints such that a node cannot perform more XOR-like NC than it can [10], [14]
$$\sum_{f_j \in \mathcal{F}, l_j \in \mathcal{L}_n^O, l_i < l_j} c^{pnc2}_{f_i,l_i,f_j,l_j} + \sum_{f_j \in \mathcal{F}, l_j \in \mathcal{L}_n^O, l_i > l_j} c^{pnc2}_{f_j,l_j,f_i,l_i} \leq y_{f_i,\bar{l}_j} \quad \forall n \in \mathcal{N}, \forall f_i \in \mathcal{F}, \forall l_j \in \mathcal{L}_n^O. \quad (32)$$

TABLE I
PHYSICAL LAYER PARAMETERS.

Parameter	Symbol	Value
SINR threshold (dB)	γ	6.4
Noise power (dBm)	N_0	-100
Path-loss exponent	η	3
Far-field crossover distance (m)	d_0	1

5) In the newly formulated problem, replace $f_j \in \mathcal{F}$ with $\overline{f}_i$ in all the constraints.

IV. NUMERICAL RESULTS

In this section, we study throughput optimization problems, i.e., JRS, JRS-BiXNC, JRS-PNC, and JRS-BiPNC problems in several wireless mesh networks. We numerically solve the problems using the optimization tool MINOS 5.51 [16]. A wireless mesh network is a multihop access network that contains a gateway node connected to the Internet, and all flows are either destined for the gateway (i.e., uplink flows) or generated by the gateway (i.e., downlink flows). The total number of flows in each network is set to be $2(N-1)$, with $N-1$ uplink flows to the gateway and $N-1$ downlink flows from the gateway. The weight of each uplink flow is 1 and the weight for each downlink flow is w, i.e., the traffic rate ratio of a downlink flow to an uplink flow is w. The physical layer parameters are given in Table I [10].

A. Bi-directional Physical Network Coding vs. Full Physical Network Coding

We use two 9-node networks (Rand9A and Rand9B) to compare the performance of bi-directional PNC and full PNC as the computational complexity for full PNC of a large size network is too large. The positions of the nodes are chosen using an uniform random distribution. The two 9-node networks are shown in Fig. 2, where the gateway node is labeled by a rectangle. We determine the weighted max-min per node throughput, i.e., $\lambda^* + w\lambda^*$ (where λ^* is the solution to the joint problem) for the two networks. In Table II, we show the weighted max-min per node throughputs for the JRS-PNC and JRS-BiPNC designs at $w = 1$ and $w = 2$ for different transmit power levels for the Rand9A and Rand9B networks. We also compute the relative throughput difference (in percentage) between the JRS-PNC and JRS-BiPNC designs as

$$\%\ Diff. = \frac{\lambda_{JRS-PNC} - \lambda_{JRS-BiPNC}}{\lambda_{JRS-PNC}} \times 100$$

where $\lambda_{JRS-PNC}$ and $\lambda_{JRS-BiPNC}$ are the weighted max-min per node throughputs for the JRS-PNC and JRS-BiPNC designs, respectively. We find that the maximum throughput difference is less than 2.5%. Thus, only a small amount of throughput is lost if the bi-directional PNC model is used instead of full PNC for the networks under consideration and under the assumption that all the uplink flows (resp. downlink flows) have the same weight. In the following, we use bi-directional PNC instead of full PNC to study wireless mesh networks.

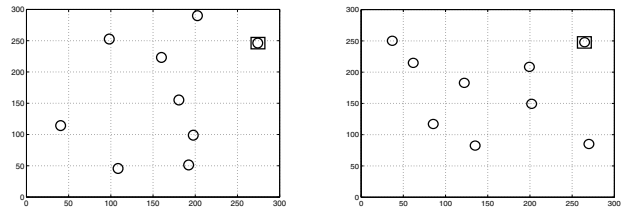


Fig. 2. The two 9-node networks: (*Left*) Rand9A and (*Right*) Rand9B.

TABLE II
BIDIRECTIONAL PNC VERSUS FULL PNC.

Network	P_t (dBm)	w	$\lambda_{JRS-PNC}$	$\lambda_{JRS-BiPNC}$	% Diff.
Rand9A	-34	1	0.0322	0.0322	0
		2	0.0175	0.0172	1.71
	-33	1	0.0344	0.0344	0
		2	0.0200	0.0196	2.04
	-32	1	0.0384	0.0384	0
		2	0.0208	0.0204	1.92
Rand9B	-34	1	0.0416	0.0416	0
		2	0.0250	0.0250	0
	-33	1	0.0435	0.0434	0.23
		2	0.0250	0.0250	0
	-32	1	0.0465	0.0465	0
		2	0.0250	0.0250	0

B. Advantages of Physical Network Coding

We compute the max-min per node throughput for the Rand9A network for different transmit power levels with $w = 1$ and $w = 2$ for the JRS, JRS-BiXNC and JRS-BiPNC designs. In Fig. 3, we present the weighted max-min per node throughputs for the different designs for different transmit power levels. The results show that, at low transmission power, PNC can provide a significant throughput gain. At higher transmission power, PNC becomes less attractive as there are more and more single-hop paths to the gateway. The throughput gain at $w = 2$ is less than the throughput gain at $w = 1$. We also find that the bidirectional XOR-type NC does not provide a significant gain for this network. One of the important advantages of PNC is that the maximum achievable throughput of a network can be obtained using a low transmit power. In Fig. 3, we find that using the JRS-BiPNC design the maximum achievable throughput is obtained with 10 dB and 6 dB lower transmit power than the other designs for $w = 1$ and $w = 2$, respectively.

To quantify the throughput gain obtained by the PNC and XOR-like NC, at low transmit power, we obtain the solution of the problems for twenty random 16-node mesh networks. One of the 16-node network is shown in Fig. 4, where the gateway node is labeled by a rectangle. We compute the weighted max-min throughput for the JRS, JRS-BiXNC, and JRS-BiPNC designs at transmit power $P_t = -26$ dBm and then calculate the throughput gain of JRS-BiPNC and JRS-BiXNC designs with respect to the JRS design. In Fig. 5, we show cumulative distribution function (CDF) of throughput gain for XOR-like NC and PNC. Clearly, the PNC provides a significant throughput gain in wireless mesh networks. The throughput gain at $w = 2$ is roughly half of the throughput gain at $w = 1$. We attribute this to the fact that the PNC opportunity

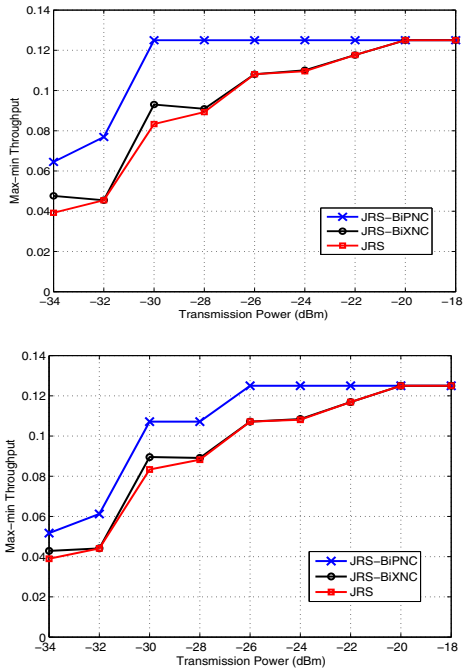


Fig. 3. Per node max-min throughput for the Rand9A network: (*Top*) $w = 1$ and (*Bottom*) $w = 2$.

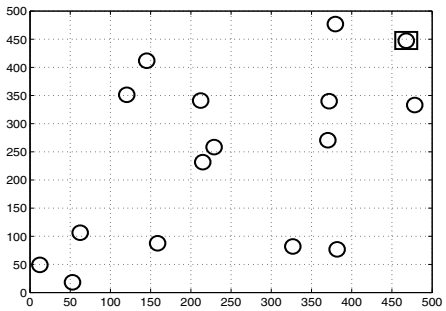


Fig. 4. The 16-node network.

is the maximum when the uplink and the downlink traffic is balanced, i.e., when $w = 1$ and the opportunity reduces with increasing the imbalance between the uplink and downlink traffic. Since the uplink traffic is half of the downlink traffic at $w = 2$, the PNC opportunity as well as the throughput gain decreases. We find that the average throughput gain of PNC is 53.67% at $w = 1$ while the gain is 30.83% at $w = 2$. We also find that the average throughput gains of XOR-like NC are 9.44% and 6.41% at $w = 1$ and $w = 2$, respectively. Thus, the throughput gain provided by the PNC is significantly higher than the throughput gain provided by the XOR-like NC in wireless mesh networks.

V. Conclusion

We have studied the joint configuration of routing, scheduling, and PNC in fixed wireless mesh networks. We have formulated the joint routing, scheduling, and PNC optimization

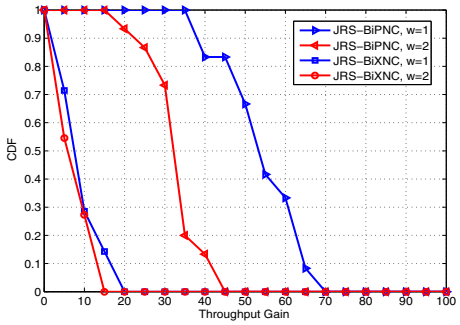


Fig. 5. Performance of XOR-like NC and PNC.

problem to maximize the minimum throughput of the flows in fixed wireless networks. The solution of the optimization problem not only provide the optimal max- min throughput but also the optimal configuration of routing parameters and the scheduling parameters with and without PNC. We have solved the problem numerically for several wireless mesh networks and demonstrated that the PNC provides a significant throughput gain in wireless mesh networks.

References

[1] R. Ahlswede, N. Cai, S. R. Li, and R. W. Yeung, "Network information flow," *IEEE Trans. Inf. Theory*, vol. 46, no. 4, pp. 1204- 1216, July 2000.

[2] C. Fragouli, J. Widmer, and J.-Y. L. Boudec, "Network coding: an instant primer," *ACM SIGCOMM Comp. Comm. Rev.*, vol. 36, pp. 63-68, Jan. 2006.

[3] D. Lun, N. Ratnakar, R. Koetter, M. Médard, E. Ahmed, and H. Lee, "Achieving minimum-cost multicast: a decentralized approach based on network coding," *Proc. IEEE INFOCOM*, pp. 1608-1617, Aug. 2005.

[4] D. S. Lun, M. Médard, and R. Koetter, "Network coding for efficient wireless unicast," *Int. Zurich Seminar on Comm.*, pp. 74-77, Feb. 2006.

[5] Y. Wu, P. A. Chou, and S.-Y. Kung, "Information exchange in wireless networks with network coding and physical-layer broadcast," *Proc. 39th Conf. on Inform. Sciences and Sys.*, Mar. 2005.

[6] S. Katti, H. Rahul, W. Hu, D. Katabi, M. Médard, and J. Crowcroft, "XORs in the air: practical wireless network coding," *Proc. ACM SIGCOMM*, pp. 243–254, Sept. 2006.

[7] Y. Wu, "Network coding for wireless networks," *Microsoft Research Technical Report MSR-TR-2007-90*, July 2007.

[8] S. Zhang, S.C. Liew, and P. Lam, "Physical Layer network coding," in *Proc. ACM Mobicom*, 2006.

[9] A. Karnik, A. Iyer, and C. Rosenberg, "Throughput- optimal configuration of fixed wireless networks," *IEEE/ACM Trans. Netw.*, vol. 16, no. 5, pp. 1161-1174, July 2007.

[10] M. F. Uddin, C. Rosenberg, W. Zhuang, P. Mitran, and A. Girard, "Joint routing and medium access control in fixed random access multi-hop networks," *IEEE/ACM Trans. Netw.*, 2013.

[11] S. Zhang, S.C. Liew, P.P. Lam, "Hot topic: physical-layer network coding", in *ACM MobiCom* , pp. 358-365, Sept. 2006.

[12] A. Iyer, C. Rosenberg, and A. Karnik, "What is the right model for wireless channel interference?" *Proc. of* 3rd *International Conference on Quality of Service in Heterogenous Wired/Wireless Networks,* 2006.

[13] S. Sengupta, S. Rayanchu, and S. Banerjee, "An analysis of wireless network coding for unicast sessions: the case for coding-aware routing," *Proc. IEEE INFOCOM*, pp. 1028-1036, 2007.

[14] S. Shabdanov, , C. Rosenberg, and P. Mitran, "Joint routing, scheduling, and network coding for wireless multihop networks" in *Proc. IEEE WiOpt*, May 2011.

[15] K. Jain, J. Padhye, V. Padmanabhan, and L. Qiu, "Impact of interference on multi-hop wireless network performance," *Proc. of ACM MobiCom'03*, pp. 66-80, Sept. 2003.

[16] Stanford Business Software, Inc., Mountain View, CA, USA. "MINOS 5.51," [Online]. Available: http://www.sbsi-sol-optimize.com/asp/sol_product_minos.htm

Unfairness Problem in WLANs due to Asymmetric Co-channel Interference and Its Mitigation

Md. Sultan Mahmud and Md. Forkan Uddin
Dept. of Electrical and Electronic Engineering
Bangladesh University of Engineering and Technology
Dhaka, Bangladesh
Email: sultan_908888@yahoo.com, mforkanuddin@eee.buet.ac.bd

***Abstract*—Co-channel interference exists in IEEE 802.11 based wireless local area networks (WLANs) due to the limited number of non-overlapping channels in the current standards. In this paper, we investigate the impact of co-channel interference in carrier sense multiple access with collision avoidance (CSMA/CA) based WLANs via simulation. We find that the throughput unfairness among the users is severe in WLANs due to asymmetric co-channel interference among the users. We propose a centralized algorithm based on analytical analysis to configure the minimum contention window of the nodes to mitigate the throughput unfairness problem in WLANs. The performance of the proposed algorithm is evaluated by simulation and found to be very effective.**

***Index Terms*—Wireless local area networks, co-channel interference, CSMA/CA, contention window, throughput, fairness.**

I. Introduction

The demand of wireless local area networks (WLANs) has grown explosively as wireless technology has evolved to meet the fundamental requirements of all categories of users. Currently, most of the WLANs are based on IEEE 802.11 standard which uses carrier sense multiple access with collision avoidance (CSMA/CA) medium access control (MAC) protocol [1]. IEEE 802.11 based WLANs use two unlicensed frequency spectrum bands: 2.4 GHz Industrial, Scientific, and Medical (ISM) band and 5 GHz Unlicensed National Information Infrastructure (UNII) band [2]. There are 14 channels available in IEEE 802.11 b/g. However, only three channels are non-overlapping among the total 14 available channels [3], [4]. Due to the limited number of non- overlapping channels, WLANs reuse the same frequency spectrum and co-channel interference exists in WLANs. Throughput is one of the important performance metrics in WLANs. Throughput performance in WLANs severely degrades due to co-channel interference [5].

WLANs suffer from two types of interferences: intra-WLAN interference and inter-WLAN interference. Intra-WLAN interference is the interference received from the stations (STAs) within the same WLAN. Whereas the inter-WLAN interference is the interference received from STAs of all the other surrounding co-channel WLANs which is known as co-channel interference. Interference is managed in WLANs using carrier sensing mechanism of CSMA/CA protocol. In CSMA/CA, a node keeps silent/idle when it senses the medium busy due to interference. The intra-WLAN interference is usually equal for all the users in a WLAN. However, co-channel interference, i.e., inter-WLAN interference is not the same to all the users of a WLAN due to the different positions of the users. Thus, the carrier sensing time of the users in a WLAN are not equal. As a result, there are throughput variations among the users. Throughput fairness is another important performance metric in wireless networks. Hence, throughput fairness among the users may severely degrade due to co-channel interference. Thus, it is very important to mitigate the throughput unfairness problem in WLANs.

In CSMA/CA based WLANs, throughputs of the users significantly depend on the contention window (CW) parameter of the users. By configuring the CW parameters of the users, throughputs of the users can be adjusted. Thus, throughput unfairness problem can be mitigated by an effective configuration of the CW parameters of the nodes. However, how to configure the CW of the nodes to mitigate the throughput unfairness problem due to the co-channel interference is not known.

In this paper, we consider single data rate WLANs to study the impact of co-channel interference. We consider a *physical interference model* based on signal to interference plus noise ratio (SINR) for the study since it is a more realistic interference model for wireless interference [6]. We consider seven co-channel WLANs where the users are randomly distributed and they use equal transmit power. Our main contributions in this paper are:

- We investigate the throughput unfairness problem due to co-channel interference in WLANs via simulation.
- Based on analytical analysis, we propose a centralized algorithm to configure the CW parameter of each user in each WLAN to mitigate the throughput unfairness problem.
- By extensive simulation campaign, we show that our algorithm is very effective to mitigate the throughput unfairness problem in WLANs.

The rest of this paper is organized as follows. We review the related work in Section II. Section III presents the network and system model in detail. In section IV, we study the impact of co-channel interference and throughput unfairness problem in WLANs. The algorithm to mitigate the throughput unfairness problem is analyzed and described in Section V.

The effectiveness of the proposed algorithm is verified in Section VI. Section VII concludes the paper.

II. Related Work

Many research work has been focused on co-channel interference in WLANs and provide different solutions to improve the performance of WLANs. In [5], Tiwary *et al.* have studied the impact co-channel interference in practical multiple access point (AP) scenario and found that the network throughput severely degrades due to co-channel interference. In [7], the authors have investigated that the network capacity significantly depends on transmit power and carrier sensing threshold. The co-channel interference increases with increasing the transmit power. However, the received signal power at a receiver is also increases with increasing transmit power. The optimal carrier sensing threshold is related to the transmit power and it is shown that for a particular data rate the product of transmit power and carrier sense threshold is constant. The authors also proposed a power control algorithm to minimize the interference. The impact of transmit power and carrier sensing threshold in WLANs is also studied in [4], [8], [9], [10]. In [2], Chieochan *et al.* suggested to minimize interference by proper placement of APs. However, the interference cannot fully mitigate using this technique. So far the research on co-channel interference of WLANs has been focused on to improve the throughput performance. To the best of our knowledge, the throughput unfairness problem due to co-channel interference is not addressed. In this paper, we investigate the throughput unfairness problem in WLANs due to co-channel interference and provide a solution to mitigate the problem.

III. System Model

A. Network Topology

Consider a WLAN which has six co-channel WLANs. We consider hexagonal structure for the co-channel WLANs with cluster size of 3. Each co-channel WLAN consists of an AP and a set of users. The AP of a WLAN is located at the center of the WLAN to provide maximum coverage. The users in the WLANs are randomly distributed and their positions are known. The total number of users in the WLANs is N. Denote the set of all the users by $\mathcal{N}$. Each node has an omni-directional antenna and cannot transmit and receive at the same time. The transmission power of all the users is fixed and equal. We consider only uplink in our model which means that only users access their APs. We also consider that the users are in saturation, i.e, they always have packets to send to their APs. We assume a single rate system where all the nodes use the same modulation and coding scheme. All the nodes use a basic CSMA/CA MAC protocol. The carrier sensing threshold (CST) for all the nodes is assumed to be the same.

B. Propagation and Interference Model

The channel gain between two nodes is assumed to be time-invariant. The channel gain between nodes n_1 and n_2, G_{n_1,n_2} is given by $(\frac{d}{d_0})^{-\eta}$, where d is the distance between the nodes, d_0 is the far-field cross-over distance and η is the path loss exponent of medium. We consider a *physical interference model* for wireless interference. Let the minimum required SINR at the receiver for successful packet decoding is γ for the given modulation and coding scheme. A packet sent by transmitter n_1 will be successfully received by the receiver n_2 if the received SINR is not less than the minimum required SINR during the reception of the packet, i.e., a packet transmission from node n_1 to n_2 will be successful if

$$\frac{G_{n_1,n_2}P_t}{N_0 + \sum_{n' \in \mathcal{N} \setminus n_1} G_{n',n_2} P_t Y_{n'}} \geq \gamma \tag{1}$$

where P_t is the transmit power, N_0 is the noise power and $Y_{n'}$ is a binary variable taking value 1 if node n' transmits and 0 otherwise.

C. Basic CSMA/CA MAC Protocol

Every node senses the medium for a time equal to the distributed interframe space (DIFS) before initiating a transmission of a packet. A node n senses medium as busy if the received power is higher than the CST, i.e., if

$$\sum_{n' \in \mathcal{N} \setminus n} P_t G_{n',n} Y_{n'} + N_0 \geq CST. \tag{2}$$

If the medium is sensed idle by a node in DIFS period it transmits its data packet. Otherwise, the node sets a value of CW randomly choosing from a uniform distribution (0, CW_{min}-1), where CW_{min} is the minimum CW of the nodes. Hence, after the DIFS time, the time is divided into mini-slots and each node senses the channel in each mini-slot if they do not transmit. If a node senses the channel as busy in a mini-slot, the node does not reduce its CW. If the channel is found as idle in a mini-slot, the node reduces its CW by one. This is called back-off process. Each node continues the back-off process until their CWs become zero. A node transmits its data packet when its CW becomes zero. Since all the nodes randomly chose the values of CWs, more than one node might have the same CW. In that case, they will start transmission at the same time and the success or failure of a transmission will depend on the used interference model. We assume that packet sizes of all the nodes are equal. If a packet transmission is successful, the receiver will send an acknowledgement (ACK) packet after a short inter frame space (SIFS) time. If a node receives an ACK packet which has already transmitted a data packet, it will attempt to send the next data packet according to the procedure mentioned above.

If a data packet transmission fails, the packet will be retransmitted. During retransmission, the value of CW is randomly chosen from a uniform distribution (0, $2^m CW_{min}$-1), where m is the number of retransmission attempts to send the packet. For example, a node will choose CW from a uniform distribution (0, $2CW_{min}$- 1) after the first failure. This is called BEB mechanism. When a data packet transmission is successful and an ACK packet is received by the sender, for the next data packet, CW will be reset by randomly choosing from a uniform distribution (0, CW_{min}-1). If a packet is not successful for five retries, the data packet is dropped.

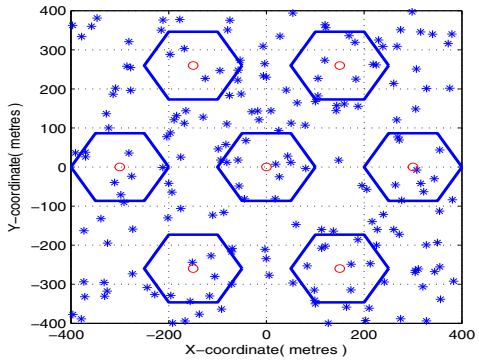


Fig. 1. The positions of the nodes in WLANs.

TABLE I
PHYSICAL LAYER PARAMETERS

Parameter	Value
SINR threshold (dB)	6.4
Noise power (dBm)	−100
Far-field crossover distance (m)	1
Data rate (Mbps)	2

IV. IMPACT OF CO-CHANNEL INTERFERENCE

In this section, we study the impact of co-channel interference in WLANs by simulation. Firstly, we study the impact of the configuration of transmit power and CST on co-channel interference and throughput performance of WLANs. Then, we study the throughput unfairness problem due to co-channel interference.

A. Network and Algorithm Parameters

We consider a WLAN and its six co-channel WLANs as shown in Fig. 1. The radius of each WLAN is 100 m. The AP of each WLAN is set at the center of the WLAN and labeled by a circle. The users are randomly distributed in $800 \times 800\ m^2$ area. Among all the users in the area, only 55 users belong to 7 co-channel WLANs including 11 users of the centre WLAN. We consider IEEE 802.11b standard for the WLANs. The physical layer parameters and MAC parameters are given in Table I and Table II, respectively. The path exponent is taken to be 3.5 considering indoor environment.

B. Simulator Setup

We simulate WLANs with a custom simulator written in C++. We have not used the commonly available network simulators as non of them provided the flexibility of tuning the network parameters. The average rates of the sources are set to equal values and their traffic is assumed to be Poisson. The source rate of the users is kept sufficiently large to make their queues saturated. To determine the total throughput for a given network scenario, the network is simulated for 500 s and the total throughput is calculated as $1058 \times 8 \times N_{pkt}/500$ bps, where N_{pkt} is the total number of packets received by the APs at the end of the simulation. For a network scenario, we have taken the average throughput over five simulation runs.

TABLE II
MAC LAYER PARAMETERS

Parameter	Value
SIFS (μs)	10
DIFS (μs)	50
a slot time (μs)	20
Data packet size (byte)	1058
ACK packet size (byte)	14
PHY + MAC header (byte)	24

C. Impact of Carrier Sensing Threshold and Transmit Power on Throughput Performance

To determine the optimal configuration of transmission power and CST for the WLAN scenario in Fig. 1, we simulate the network by varying transmit power of the users at different CST power levels. The minimum CW of the nodes is set to 31 which is the default minimum CW of the nodes. In Fig. 2 *(Top)*, we show the total throughput of the WLANs with respect to the transmit power for different CSTs. For a given CST, the throughput performance is poor at a low transmit power due to the hidden terminal problem[1]. The number of hidden terminals reduces with increasing the transmission power and hence the total throughput increases with increasing the transmit power from a low transmit power. Further, the total throughput starts to decrease with increasing the transmit power from the optimal transmit power for decreasing the spatial reuse due to increasing the co-channel interference. At a high transmit power, the throughput performance is poor since the spatial reuse reduces significantly due to severe co-channel interference. We find that the optimal value of the transmit power increases with increasing the CST power. We also find that the total throughput is not good at a low CST even when the transmit power is optimized. Interestingly, the throughputs for CST power levels -96, -94, and -92 dBm with the corresponding optimal transmit power are very close. The noise power is set to -100 dBm for simulation. Thus, the CST should be roughly 4 to 8 dB higher than the noise power to obtain a better throughput performance with the optimal transmit power.

For the system in Fig. 1, if the distance between two transmitting nodes is higher than 200m, they are not hidden terminals. To avoid hidden terminals completely, i.e., assuming 200 m distance as carrier sensing range, the transmit power levels are found to be -23.39, -19.27, -16.32 and -13.81 dBm for CST of -98, -96, -94 and -92 dBm, respectively. Our simulation results show that the throughputs obtained by these transmit power levels are very close to the optimal throughput. Thus, one can obtain a very close throughput to the optimal throughput by setting the transmit power to a value such that the hidden terminal problem is avoided completely.

D. Impact of Contention Window on Throughput Performance

In [11], G. Bianchi analyzes throughput performance in single WLAN. The author provides the expression of the

[1]A hidden terminal is a terminal which does not sense carrier for a transmission by a transmitter to a receiver but a transmission by the hidden terminal causes collision to the receiver.

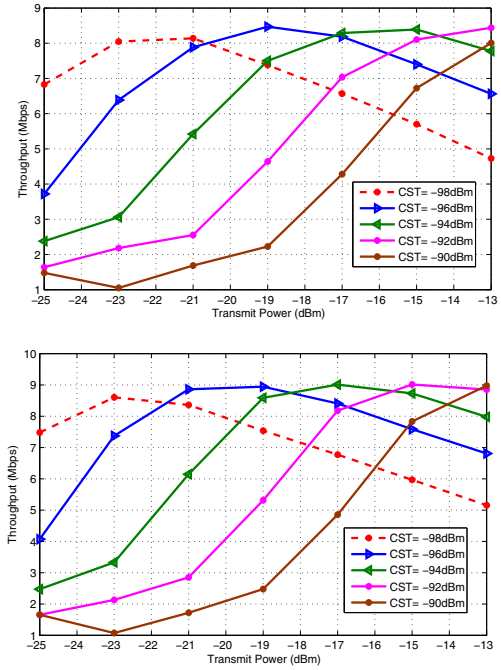


Fig. 2. The total network throughput with respect to transmit power: *(Top)* Default contention window and *(Bottom)* Bianchi optimal contention window.

optimal value of the minimum CW as $N_c\sqrt{2T_c}$ to achieve the maximum throughput, where N_c is the number of contending nodes in the WLAN and T_c is the collision time in mini-slots. For our WLAN system, the collision time is found to be 219 mini-slots using the following equation.

$$T_c = \frac{8(L_d + PHY + MAC)}{C\delta} + \frac{DIFS}{\delta} \tag{3}$$

where L_d is the data packet size, PHY and MAC are the physical and MAC layer headers, respectively, and δ is the duration of a mini-slot. We calculate the Bianchi optimal CW of the nodes for each WLAN of the system shown in Fig. 1. Using Bianchi optimal CW, we determine the total throughput of the system by varying the transmit power under different CSTs. The throughput performance at Bianchi optimal CW is shown in Fig. 2 *(Bottom)*. We find that the optimal throughput is 9.01 Mbps for Bianchi optimal CW while the optimal throughput is 8.5 Mbps for default CW. The results show that the insights on the impact of transmit power and the CST remain the same even when the Bianchi optimal CW is used.

E. Unfairness Problem due to Co-channel Interference

The throughput of each user of the centre WLAN in Fig. 1 is determined at $P_t = -15$ dBm and CST=-92 dBm with the default as well as Bianchi optimal CW. The throughput of each user of the centre WLAN is shown in Fig. 3, where user index is labeled in the x-axis. The minimum and the maximum throughputs of the users are found to be 0.04 Mbps and 0.174 Mbps, respectively, at default CW. Thus, the throughput unfairness is very significant at default CW configuration of

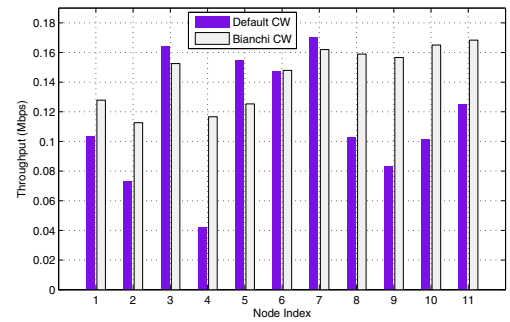


Fig. 3. Throughput of each user of the centre WLAN.

the nodes. On the other hand, the minimum and the maximum throughputs of the users are found to be 0.114 Mbps and 0.173 Mbps, respectively, at Bianchi optimal CW. Though configuring the WLANs with Bianchi optimal CW can reduce the throughput unfairness problem, the unfairness problem is still significant.

V. Algorithm for Solving the Throughput Unfairness Problem

In this section, we propose an algorithm to configure the minimum CW of the nodes to solve the throughput unfairness problem based on an analytical analysis. Consider WLANs composed of several co-channel WLANs. We assume that node i belongs to one of the WLANs. Each node in a CSMA/CA based wireless network passes time mainly in four states: idle, carrier sense, successful transmission and collision. In the following, we derive the expressions of fraction of times in different states of node i, relate the fraction of times of different states, and describe the algorithm to configure the minimum CW of the nodes.

A. Fraction of Times in Different States

We divide the carrier sensing time into three parts: carrier sensing for transmission of one node only, carrier sensing for transmissions of two nodes simultaneously, and carrier sensing for transmission of three nodes simultaneously. The carrier sensing time for transmission of more than three nodes at the same time is very less and it is neglected in the analysis. Let Π_i^{idle}, Π_i^{cs}, Π_i^{cs-2}, Π_i^{cs-3}, Π_i^{s} and Π_i^{c} be the fraction of times that the node i passes times in idle state, carrier sensing state for transmission of single node, carrier sensing state for transmission of two nodes, carrier sensing state for transmission of three nodes, successful state, and collision state, respectively.

We want to provide max-min throughput fairness among the users. Let λ packet/s be the max-min throughput of each user. The fraction of successful time can be written as

$$\Pi_i^s = k\lambda\delta \tag{4}$$

where, k is the transmission time in mini-slots for a packet.

The total fraction of time in successful and collision states can be written as

$$\Pi_i^s + \Pi_i^c = \frac{k\lambda\delta}{1-p_i} \tag{5}$$

where p_i is the collision probability of a transmission of node i. The fraction of time in idle state can be written as

$$\begin{aligned}\Pi_i^{idle} &= \frac{CW_{min}^i}{2}\delta\lambda(1+2p_i+2^2p_i^2+\ldots+2^mp_i^m)\\ &= \frac{CW_{min}^i}{2}\delta\lambda\left[\frac{1-(2p_i)^{m+1}}{1-2p_i}\right]\end{aligned} \tag{6}$$

where CW_{min}^i is the minimum CW of node i. Assuming that the nodes access the medium independently when they are in idle state, the fraction of times in the carrier sensing states can be written as

$$\Pi_i^{cs} = \sum_{n\in\mathcal{N}_i^{cs}} \frac{k\lambda\delta}{1-p_n} \tag{7}$$

$$\Pi_i^{cs-2} = \sum_{(x,y)\in\mathcal{N}_i^{cs-2}} \frac{k\lambda\delta}{(1-p_x)}.\frac{k\lambda\delta}{(1-p_y)} \tag{8}$$

$$\Pi_i^{cs-3} = \sum_{(x,y,z)\in\mathcal{N}_i^{cs-3}} \frac{k\lambda\delta}{(1-p_x)}.\frac{k\lambda\delta}{(1-p_y)}.\frac{k\lambda\delta}{(1-p_z)} \tag{9}$$

where $\mathcal{N}_i^{cs}$ is the set of nodes such that node i senses carrier if one of the nodes in $\mathcal{N}_i^{cs}$ transmits, $\mathcal{N}_i^{cs-2}$ is the set of pair of nodes such that node i senses carrier if one of the pair of nodes in $\mathcal{N}_i^{cs-2}$ transmit simultaneously, and $\mathcal{N}_i^{cs-3}$ is the set of triplet of nodes such that node i senses carrier if one of the triplet of nodes in $\mathcal{N}_i^{cs-3}$ transmit simultaneously. Note that $\mathcal{N}_i^{cs-2}$ does not include any node in $\mathcal{N}_i^{cs}$ and $\mathcal{N}_i^{cs-3}$ does not include any node or pair of nodes in $\mathcal{N}_i^{cs}$ and $\mathcal{N}_i^{cs-2}$. Assuming that each node shares the medium equally in a WLAN, we approximate $\frac{k\lambda\delta}{(1-p_x)}$ as $\frac{1}{N_x}$, where N_x is the number of contending nodes in the WLAN of node x. Thus, (8) and (9) can be approximated as

$$\Pi_i^{cs-2} \approx \sum_{(x,y)\in\mathcal{N}_i^{cs-2}} \frac{1}{N_x}.\frac{k\lambda\delta}{(1-p_y)} \tag{10}$$

$$\Pi_i^{cs-3} \approx \sum_{(x,y,z)\in\mathcal{N}_i^{cs-3}} \frac{1}{N_x}.\frac{1}{N_y}.\frac{k\lambda\delta}{(1-p_z)}. \tag{11}$$

B. Relation Among the Fraction of Times

The summation of the fraction of times of all the states of node i is one. Thus,

$$\begin{aligned}&\frac{CW_{min}^i}{2}\delta\lambda\left[\frac{1-(2p_i)^{m+1}}{1-2p_i}\right] + \sum_{n\in\mathcal{N}_i^{cs}}\frac{k\lambda\delta}{1-p_n}\\ &+ \sum_{(x,y)\in\mathcal{N}_i^{cs-2}}\frac{1}{N_x}.\frac{k\lambda\delta}{(1-p_y)} + \sum_{(x,y,z)\in\mathcal{N}_i^{cs-3}}\frac{1}{N_x}.\frac{1}{N_y}.\frac{k\lambda\delta}{(1-p_z)}\\ &+ \frac{k\lambda\delta}{1-p_i} = 1.\end{aligned} \tag{12}$$

The set $\mathcal{N}_i^{cs}$ includes all the nodes of self WLAN of node i and some nodes from the co-channel WLANs of node i. Denote by $\mathcal{N}_i^{cs-s}$ and $\mathcal{N}_i^{cs-o}$ the sets of nodes from self WLAN and co-channel WLANs, respectively, such that $\mathcal{N}_i^{cs-s}\cup\mathcal{N}_i^{cs-o} = \mathcal{N}_i^{cs}$. Assuming that there is no hidden terminal in the system, i.e., a node senses carrier if one of the nodes from the same WLAN transmits and $(2p_i)^{m+1} = 0$, (12) can be written as

$$\begin{aligned}&\frac{CW_{min}^i}{2}\delta\lambda\left[\frac{1}{1-2p_i}\right] + \sum_{n\in\mathcal{N}_i^{cs-o}}\frac{k\lambda\delta}{1-p_n}\\ &+ \sum_{(x,y)\in\mathcal{N}_i^{cs-2}}\frac{1}{N_x}.\frac{k\lambda\delta}{(1-p_y)} + \sum_{(x,y,z)\in\mathcal{N}_i^{cs-3}}\frac{1}{N_x}.\frac{1}{N_y}.\frac{k\lambda\delta}{(1-p_z)}\\ &+ \sum_{n\in\mathcal{N}_i}\frac{k\lambda\delta}{1-p_n} = 1\end{aligned} \tag{13}$$

where $\mathcal{N}_i$ is the set of nodes in the WLAN of node i,.i.e., $\mathcal{N}_i = \mathcal{N}_i^{cs-s}\cup i$.

C. Algorithm

Assume an arbitrary node j in the same WLAN of node i which does not sense carrier for transmissions of the nodes in co-channel WLANs. Thus, for node j (13) can be written as

$$\frac{CW_{min}^j}{2}\delta\lambda\left[\frac{1}{1-2p_j}\right] + \sum_{n\in\mathcal{N}_j}\frac{k\lambda\delta}{1-p_n} = 1. \tag{14}$$

Since $\mathcal{N}_i = \mathcal{N}_j$, from (13) and (14) we obtain,

$$\begin{aligned}&\frac{CW_{min}^i}{2}\left[\frac{1}{1-2p_i}\right] + \sum_{n\in\mathcal{N}_i^{cs-o}}\frac{k}{1-p_n}\\ &+ \sum_{(x,y)\in\mathcal{N}_i^{cs-2}}\frac{1}{N_x}.\frac{k}{(1-p_y)} + \sum_{(x,y,z)\in\mathcal{N}_i^{cs-3}}\frac{1}{N_x}.\frac{1}{N_y}.\frac{k}{(1-p_z)}\\ &= \frac{CW_{min}^j}{2}\left[\frac{1}{1-2p_j}\right].\end{aligned} \tag{15}$$

Since the users of the co-channel WLANs are located far in distance from the AP (receiver) of the WLAN of node i, the collision probability for a transmission by node i can be written as

$$p_i = 1 - \prod_{n\in\mathcal{N}_i^{cs-s}}(1-\tau_n) = 1 - \frac{D}{1-\tau_i} \tag{16}$$

where τ_i is the attempt rate of node i in each mini-slot of idle state and $D = \prod_{n\in\mathcal{N}_i}(1-\tau_n)$. The variable τ_i can be approximated as [11]

$$\tau_i \approx \frac{2}{CW_{min}^i}. \tag{17}$$

Using (17), the collision probability of a node of a co-channel WLAN is approximated as follows using the Bianchi optimal attempt rate [11].

$$p_n = 1-(1-\tau)^{N_n-1} = 1-\left(1-\frac{2}{N_n\sqrt{2T_c}}\right)^{N_n-1}. \tag{18}$$

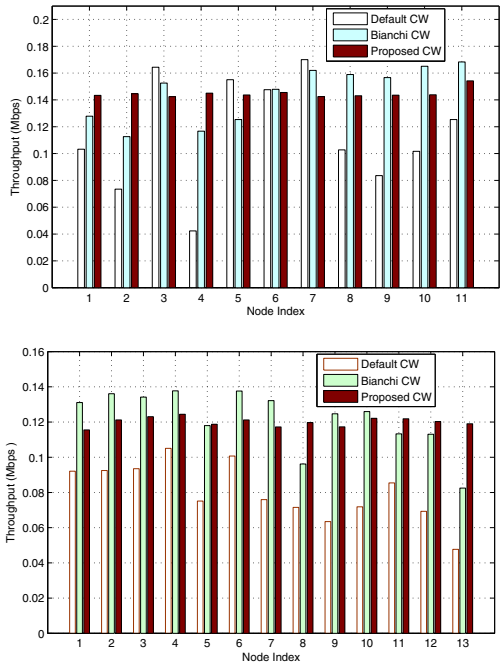


Fig. 4. Throughput of each user of the centre WLAN: *(Top)* WLAN scenario-1 (Fig. 1) and *(Bottom)* WLAN scenario-2.

Using (16), (17), and (18) in (15), we get a relation between CW^i_{min} and CW^j_{min} in term of D. Hence, the other parameters in (15) become constant for a given WLAN scenario. We set Bianchi optimal CW for CW^j_{min}, i.e., $CW^j_{min} = N_j\sqrt{2Tc}$ and then determine the value of CW^i_{min} for all i in the WLANs under a suitable value of D.

VI. Effectiveness of the Proposed Algorithm

We configure the minimum CW of the nodes for the WLAN scenario in Fig. 1 according to the algorithm in Section V at $P_t = -15$ dBm and $CST = -92$ dBm and determine the throughput of the WLANs by simulation. We use the value of $k = 215$ in computation of the minimum CW of the nodes since $k = (L_d + SIFS + ACK)/\delta$. In Fig. 4 *(Top)*, we show the throughput of each user of the centre WLAN under three types of configurations of the CW of the nodes. The results show that the minimum throughput of the users is 0.142 Mbps and the maximum throughput of the users is 0.152 Mbps when the CW of the nodes are configured using our algorithm. Thus, our algorithm is very effective to solve the throughput unfairness problem in WLANs due to co-channel interference.

We know that there is a trade-off between the total throughput and fairness in wireless networks. If one wants to provide max-min fair throughput among the users then the total throughput degrades significantly. To verify this, we determine the total throughput of the WLANs under Bianchi optimal and our proposed configurations of WLANs. We find that the total throughput under Bianchi optimal CW is 9.013 Mbps while the total throughput under our algorithm is 9.011 Mbps. Thus, the throughput lost is not significant to provide the max-min throughput fairness among the users with our algorithm.

We also determine the throughput of each user of the centre WLAN for an another WLAN scenario similar to the WLAN scenario in Fig. 1 under 3 types of configurations of CW of the nodes. The throughput of the users of the centre WLAN is shown in Fig. 4 *(Bottom)*. The maximum and the minimum throughputs of the users of centre WLAN are found to be 0.122 Mbps and 0.118 Mbps when our algorithm is used. The total throughputs under Bianchi optimal CW and our algorithm are 8.34 Mbps and 8.33 Mbps, respectively. Thus, the results demonstrate that our algorithm is very effective to provide throughput fairness among the users without losing total throughput considerably.

VII. Conclusion

We have studied the impact of co-channel interference in WLANs. We find that there is a severe throughput unfairness among the users due to co-channel interference. We propose a centralized algorithm to configure the minimum contention window of the nodes to provide throughput fairness among the users. Our simulation results show that our algorithm is very effective to provide fairness among the users. The future work related to this work is to develop a distributed algorithm to configure the minimum contention window of the nodes to provide throughput fairness among the users in WLANs.

References

[1] IEEE 802.11 Working Group, "Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specification," 1997.

[2] S. Chieochan, E. Hossain, and J. Diamond,"Channel assignment schemes for infrastructure-based 802.11 WLANs: A survey," *IEEE Communications survey tutorial*, vol. 12, no. 1, First Quarter 2010.

[3] M. Haidar, R. Ghimire, H. Al-Rizzo, R. Akl and Y. Chan, "Channel assignment in an IEEE 802.11 WLAN based on signal-to- interference ratio," in *Proc. CCECE*, pp. 1169-1173, May-2008.

[4] A. Vasan, R. Ramjee, and T. Woo,"ECHOS: Enhanced Capacity 802.11 Hotspots," in *Proc. IEEE INFOCOM*, 2005.

[5] P. K. Tiwary, N. Maskey, S. Khakurel and G. Sachdeva, "Effects of co-channel interference in WLAN and cognitive radio based approach to minimize it", in *Proc. ICARTCC*, pp.158-160, 2010.

[6] A. Iyer, C. Rosenberg and A. Karnik, "What is the right model for wireless channel interference?" *IEEE/ACM Trans. on Networking*, vol. 8, no. 5, May 2009.

[7] T. Kim, H. Lim and J. C. Hou,"Improving spatial reuse through tuning transmit power, carrier sense threshold, and data rate in multihop wireless networks," in *Proc. MobiCom*, pp. 23-26, Sept. 2006.

[8] V. P. Mhatre and K. Papagiannaki,"Optimal design of high density 802.11 WLANs," in *Proc. ACM CoNEXT*, 2006.

[9] J. A. Fuemmeler, N. H. Vaidya and V. V. Veeravalli, " Selecting transmit power and carrier sense threshold for CSMA protocols," *Technical Report*, Oct. 2004.

[10] H. Zhai and Y. Fang,"Physical carrier sensing and spatial reuse in multirate and multihop wireless ad hoc networks," in *Proc. of IEEE INFOCOM,* 2006.

[11] G. Bianchi, "Performance analysis of the IEEE 802.11 distributed coordination function," *IEEE Journal On Selected Areas in Communications*, vol. 18, no. 3, March 2000.

Performance Analysis of MC-DS-CDMA in the presence of Carrier Frequency Offset and Timing Jitter over Rayleigh Fading Channels

Md. Ishtiaque Aziz Zahed[1, 2*] and S. P. Majumder[1]

[1]Department of Electrical & Electronic Engineering, Bangladesh University of Engineering & Technology
Dhaka-1000, Bangladesh

[2]Department of Electrical & Electronic Engineering, East Delta University
Chittagong-4000, Bangladesh

[*]ishtiaque2307@gmail.com

Abstract — **In this paper an analytical approach is presented to determine the impact of frequency offset, timing jitter and additive white Gaussian noise (AWGN) on the bit error rate (BER) performance of a multi-carrier direct-sequence code division multiple access (MC-DS-CDMA) system over a Rayleigh Fading Channel. The analysis developed the pdf (probability density function) at the receiver considering combined influence of fading, timing jitter and Doppler frequency offset etc with maximal ratio combining (MRC) scheme. The expression for the conditional BER conditioned on a given timing error and fading is derived and the average BER is evaluated in the presence of Multiple Access Interference (MAI) and Inter-Carrier Interference (ICI). The performance results are evaluated numerically in terms of SINR and BER considering system parameters like number of users, number of sub-carriers. The result shows significant deterioration in SINR and BER performance due to fading along with the changes in parameters.**

Index Terms — **BER, MAI, MC-DS-CDMA, Rayleigh Fading, SINR, Timing Jitter**

I. Introduction

In spread spectrum techniques correlation properties of code sequence allows the bit error rate performance to be varied; as it determines the amount of interference from multiple access known as Multiple Access Interference (MAI) and causes degradation in the quality of the received signal. Frequency selective fading also introduces interference which severely affects CDMA system. On the other hand, Orthogonal Frequency-Division Multiplexing (OFDM) is effective to avoid intersymbol interference due to multipath delay. However, it is sensitive to timing jitter caused by time synchronization error at the receiver. The effect of timing jitter, on the performance of an OFDM system without considering multipath fading is also reported [6] - [7]. Introduction of Multicarrier OFDM technique with CDMA makes the system robust to frequency selectivity, helps the system utilizing the advantage of frequency diversity by spreading the signal across multiple subcarriers. Currently, different combinations of OFDM and CDMA have been investigated and presented by various researchers [1–3]. One of these combinations is the multicarrier direct-sequence code division multiple access (MC-DSCDMA), which has been considered as an important technique for the future generation wireless systems [4].

In the MC-DSCDMA system, the transmitter spreads the serial-to-parallel converted data streams using the same spreading sequence in the time domain and transmits the parallel data symbols corresponding to the same spreading chip by OFDM modulation. The system is more sensitive to timing jitter which is due to the mismatch between actual sampling time and the ideal sampling time than single-carrier systems as a large number of subcarriers are utilized [5, 6]. Timing jitter degrades performance of the MC-DS-CDMA system because it destroys orthogonality among subcarriers and results in inter-carrier interference (ICI). It has been claimed that the performance degradation for the MC-DS-CDMA system caused by timing jitter is independent of the number of subcarriers, of the spreading factor and of the spectral contents of the jitter, but only depends on the timing jitter variance [9]. However, the performance measure employed in [9] is the signal power to interference and noise power ratio, which is an indirect indicator of system bit error rate (BER) performance.

In this paper, the analytical expression is formulated for the MC-DS-CDMA system to determine the impact of timing jitter variance and frequency offset, working over a Rayleigh fading channel for determining the expression of multiple access interference (MAI) and the expression of the SINR and BER have been deduced to observe how it affects the number of subcarriers and users.

The paper is organized as follows. The system model is described in Section II. Section III provides combined analytical expressions of BER performance for MC-DS-CDMA system in Rayleigh fading environment at the presence of frequency offset and timing jitter. Section IV gives the results of system performance analysis and finally Section V finishes the paper by giving conclusion.

II. System Model

For the analysis the system model is considered to have the following characteristics:

The users are active all time and transmit signal power at the same level. Fading for all transmitted signals received by an antenna is assumed equal. As Rayleigh fading channel is considered, the amplitude distortion coefficients are Gaussian distributed while phase shift coefficient is uniformly distributed. The input signal, only having values of +1 and -1 is transmitted using OFDM multicarrier modulation technique over a dispersive Rayleigh fading channel. Maximal ratio combining is used to combine all the outputs.

The system model considered for this analysis is shown in Fig.1 and Fig.2 where Fig.1 and Fig.2 are the block diagrams of MC-DS-CDMA transmitter and receiver respectively.

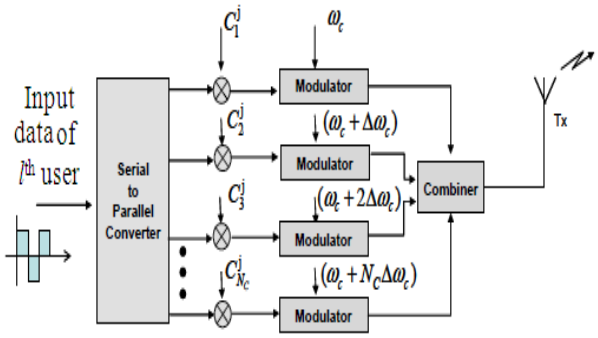


Fig. 1 Block Diagram of MC DS CDMA Transmitter

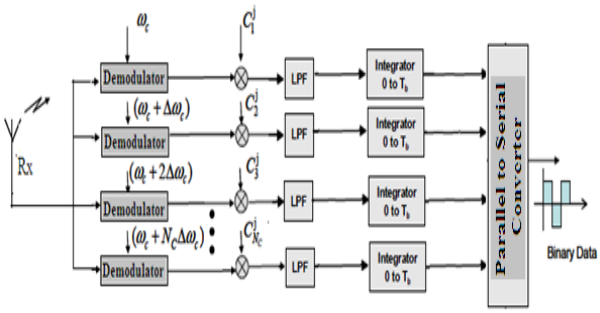


Fig. 3 Block Diagram of MC DS CDMA Receiver

Let, there are N_u-number of users and l-th user is taken as the reference for any period of time. Now, the input data of the l -th user are converted into N_c parallel data streams and each of the these parallel sequences is multiplied by a particular PN code sequence and thereby spread in the time domain. Thus each data bit is split in time domain and then is modulated by the respective sub-carrier. To write the general expression of the sub-carriers, let us consider; ω_c = Freq of the reference channel, $\Delta\omega_c$= Frequency spacing between two successive channels and φ_k = Instantaneous phase angle of the k-th sub-carrier.

The transmitted MC-DS-CDMA signal of user j; corresponding to the m-th chip of the spreading sequence can be written as

$$s_{l,m}(t) = \frac{\sqrt{2p}}{N_C}\sum_{k=0}^{N_C-1} b_{k,l}\, c_{m,l} e^{j2\pi f_k\{t-(m-1)(T_S+t_g)\}} \dots\dots(1)$$

Here, for the duration $(m-1)(T_S+t_g)-t_g \le t \le m(T_S+t_g)-t_g$,, where $f_k = k \times f_s$ is the frequency of the k-th subcarrier, $f_s = \frac{1}{T_S}$ is the subcarrier spacing related to the chip duration T_S, T_g is the guard interval, N_c denotes number of subcarrier, N_u is the number of users and $b_{k,l}$ is the data symbol transmitted by user l on the k-th subcarrier. $c_{m,l}$ is the m-th chip of data sequence that spreads the symbols from user j and L is the length of spreading sequence.

The signal for l -th user will be,

$$s_l(t) = \frac{\sqrt{2p}}{N_C}\sum_{m=1}^{L}\sum_{k=0}^{N_C-1} b_{k,l}\, c_{m,l}\, e^{j2\pi f_k\{t-(m-1)(T_S+t_g)\}} \dots\dots(2)$$

Here, the spreading sequence $c_{m,l}$ does not depend on the subcarrier index k, i.e., all data symbols from user l that are transmitted during the same symbol interval are spread with the same spreading sequence. Typically, orthogonal spreading sequences of length L are used to maintain orthogonality between signals of different users.

Then the MC-DS-CDMA signal is transmitted through a multipath Rayleigh fading wireless channel which introduces inter-symbol interference (ISI) due to the differences between time delays of multiple propagation paths. The maximum delay difference is known as the delay spread of the channel and it is the channel parameter that determines the duration of ISI. Hence, by using a guard interval t_g that is no less than the delay spread, the ISI is completely confined within the guard interval. By simply discarding the signal received over the guard interval, the receiver can get rid of the ISI.

The received signal for m-th chip can be considered as –

$$r_m(t) = \frac{\sqrt{2p}}{N_C}\sum_{l=1}^{N_u}\sum_{k=0}^{N_C-1} b_{k,l}\, c_{m,l}\alpha_{k,l} e^{j2\pi f_k\{t-(m-1)(T_S+t_g)\}} + n_{ml}(t) \dots\dots(3)$$

Where, $\alpha_{k,l}$ is the fading channel gain for the k- th subcarrier of user l and it is a complex Gaussian random process with zero mean and variance $\sigma_{k,l}^2$. Thus, each subcarrier observes a frequency-flat fading channel although the actual channel is multipath frequency-selective.

The complex-valued additive white Gaussian noise (AWGN) process $n_{ml}(t)$ has zero mean and varianceσ_N^2.

The signal received for a whole sequence

$$r(t) = \frac{\sqrt{2p}}{N_C}\sum_{m=1}^{L}\sum_{l=1}^{N_u}\sum_{k=0}^{N_C-1} b_{k,l}\, c_{m,l}\alpha_{k,l} e^{j2\pi f_k\{t-(m-1)(T_S+t_g)\}} + n_{ml}(t) \dots\dots(4)$$

As the focus is on the ICI effect, the normalized carrier frequency offset Δf_k is considered which takes the value of an integer. We consider relative CFO to be a Gaussian process, statistically independent of the input signal and the signal will be

$$r(t) = \frac{\sqrt{2p}}{N_C}\sum_{m=1}^{L}\sum_{l=1}^{N_u}\sum_{k=0}^{N_C-1} b_{k,l}\, c_{m,l}\alpha_{k,l} e^{j2\pi(f_k+\Delta f_k)\{t-(m-1)(T_S+t_g)\}} + n_{ml}(t) \dots\dots(5)$$

The complex valued additive white Gaussian noise (AWGN) process $n_{ml}(t)$ has zero mean and varianceσ_l^2. Due to sampling, $t_{n,m}$ is supposed to be $nT + (m-1)(T_S+t_g)$ But, random errors called timing jitters, occur at the receiver due to synchronization errors and limitations of sampling devices. This means the sampling instants are $t_{n,m} = nT + (m-1)(T_S+t_g) + \epsilon_{n,m}$ and $\epsilon_{n,m} \equiv T\epsilon(n,m)$, is the timing error due to nth sampling instant for MC DS CDMA. $\epsilon(n,m)$ is the timing jitter normalized by the sampling interval T, which is a wide sense stationary Gaussian process with zero mean and variance σ_J^2 .

Therefore, the actual signal samples can be expressed as

$$r_n = r(nT + (m-1)(T_S+t_g) + \epsilon_{n,m})$$

$$r_n = \frac{\sqrt{2p}}{N_C}\sum_{m=1}^{L}\sum_{l=1}^{N_u}\sum_{k=0}^{N_C-1} b_{k,l}\, c_{m,l} \times e^{j2\pi(f_k+\Delta f_k)\{nT+(m-1)(T_S+t_g)+\epsilon_{n,m}-(m-1)(T_S+t_g)\}} + n_{l,n} \ldots (6)$$

Where, $n_{l,n} = n_{ml}(nT + (m-1)(T_S + t_g) + \epsilon_{n,m})$

$$r_n = \frac{\sqrt{2p}}{N_C}\sum_{m=1}^{L}\sum_{l=1}^{N_u}\sum_{k=0}^{N_C-1} b_{k,l}\, c_{m,l}\alpha_{k,l} e^{j2\pi(f_k+\Delta f_k)\{nT+T\epsilon(n,m)\}} + n_{l,n} \ldots\ldots (7)$$

$$r_n = \frac{\sqrt{2p}}{N_C}\sum_{m=1}^{L}\sum_{l=1}^{N_u}\sum_{k=0}^{N_C-1} b_{k,l}\, c_{m,l}\alpha_{k,l} e^{j2\pi\frac{k}{NT}\{nT+T\epsilon(n,m)\}} \times e^{j2\pi\Delta f_k\{nT+T\epsilon(n,m)\}} + n_{l,n} \ldots\ldots (8)$$

$$r_n = \frac{\sqrt{2p}}{N_C}\sum_{m=1}^{L}\sum_{l=1}^{N_u}\sum_{k=0}^{N_C-1} b_{k,l}\, c_{m,l}\alpha_{k,l} e^{j2\pi\frac{k}{N}\{n+\epsilon(n,m)\}} \times e^{j2\pi\Delta f_k T\{n+\epsilon(n,m)\}} + n_{l,n} \ldots\ldots (9)$$

Typically the jitter variance σ_J^2 should not exceed 10^{-2} for the MC-DS-CDMA system to work in a suitable operating condition as otherwise the degradation would be unacceptably large [9].

III. Theoretical Analysis

The test decision variable Z_i for the data symbol carried by the *i*-th subcarrier of user *l* can be obtained from the received signal samples after despreading as

$$Z_i = \frac{\alpha_{i,l}^*}{N_C L}\sum_{n=0}^{N_C-1}\left(r_n C_{m,l}^*\right) e^{-j2\pi f_i nT} \ldots (10)$$

$$= \frac{\alpha_{i,l}^*}{N_C L}\sum_{n=0}^{N_C-1}\sqrt{2p}\sum_{m=1}^{L}\sum_{l=1}^{N_u}\sum_{k=0}^{N_C-1} b_{k,l}\, C_{m,l}C_{m,l}^*\alpha_{k,l} \times e^{j2\pi\frac{k}{N}\{n+\epsilon(n,m)\}} e^{j2\pi\Delta f_k T\{n+\epsilon(n,m)\}} e^{-j2\pi f_i nT} + n_i$$

$$= \frac{\sqrt{2p}\alpha_{i,l}^*}{N_C L}\sum_{l=1}^{N_u}\sum_{k=0}^{N_C-1} b_{k,l}\,\alpha_{k,l}\sum_{m=1}^{L} C_{m,l}C_{m,l}^* \sum_{n=0}^{N_C-1} e^{j2\pi\frac{k}{N}\{n+\epsilon(n,m)\}} e^{j2\pi\Delta f_k T\{n+\epsilon(n,m)\}} e^{-j2\pi\frac{in}{N}} + n_i$$

$$Z_i = \frac{\sqrt{2p}\alpha_{i,l}^*}{N_C L}\sum_{l=1}^{N_u}\sum_{k=0}^{N_C-1} b_{k,l}\,\alpha_{k,l}\sum_{m=1}^{L} C_{m,l}C_{m,l}^* \sum_{n=0}^{N_C-1} e^{j2\pi\frac{1}{N}\{(k-i)n+k\epsilon(n,m)\}} e^{j2\pi\Delta f_k T\{n+\epsilon(n,m)\}} + n_i \ldots\ldots (11)$$

Where the noise term,

$$n_i = \frac{\alpha_{i,l}^*}{NL}\sum_{m=1}^{L}\sum_{n=0}^{N-1}\left(n_{l,n}C_{m,l}^*\right) e^{-j2\pi f_i nT}$$

is due to the AWGN. To recognize different types of interference we break the decision variable into different components .

If we take $l=l$ and k = i, then the desired signal for l^{th} user and i^{th} subcarrier will be;

$$s_i = \frac{\sqrt{2p}}{N_C L} b_{i,l}\alpha_{i,l}^2 \sum_{m=1}^{L}\sum_{n=0}^{N_C-1} e^{j2\pi\frac{i}{N}\epsilon(n,m)\}} \ldots\ldots (12)$$

Again, for self interference, $l=l$ and k ≠ i,

$$\alpha_i = \frac{\sqrt{2p}\alpha_{i,l}^*}{N_C L}\sum_{\substack{k=0\\k\neq i}}^{N_C-1} b_{k,l}\,\alpha_{k,l}\sum_{m=1}^{L}\sum_{n=0}^{N_C-1} e^{j2\pi\frac{1}{N}\{(k-i)n+k\epsilon(n,m)\}} \times e^{j2\pi\Delta f_k T\{n+\epsilon(n,m)\}} \ldots\ldots (13)$$

Interference from other carriers of other users, we take $l\neq l$ and k ≠ i,

$$\beta_i = \frac{\sqrt{2p}\alpha_{i,l}^*}{N_C L}\sum_{\substack{l=1\\l\neq l}}^{N_u}\sum_{\substack{k=0\\k\neq i}}^{N_C-1} b_{k,l}\alpha_{k,l}\sum_{m=1}^{L} C_{m,l}C_{m,l}^* \sum_{n=0}^{N_C-1} e^{j2\pi\frac{1}{N}\{(k-i)n+k\epsilon(n,m)\}} e^{j2\pi\Delta f_k T\{n+\epsilon(n,m)\}} \ldots\ldots (14)$$

Interference from same carrier of other users, we take $l\neq l$ and k=i,

$$\mu_i = \frac{\sqrt{2p}\alpha_{i,l}^*}{N_C L}\sum_{\substack{l=1\\l\neq l}}^{N_u} b_{k,l}\,\alpha_{k,l}\sum_{m=1}^{L} C_{m,l}C_{m,l}^* \sum_{n=0}^{N_C-1} e^{j2\pi\frac{1}{N}\{(k-i)n+k\epsilon(n,m)\}} e^{j2\pi\Delta f_i T\{n+\epsilon(n,m)\}} \ldots\ldots (15)$$

From the equations it is clear that, in spite of using orthogonal codes, timing jitter and CFO have introduced not only self interference but also multi user interference. As, $b_{k,l}$'s have zero mean, so the interference components are random variables with zero mean.

Now, we can consider; signal power

$$P_s = 2P\left(\sum_{i=0}^{N_C-1}\left|\alpha_{i,l}\right|^2\right)^2 \ldots\ldots (16)$$

Variance of self interference,

$$\sigma_{CI}^2 = \frac{2P\mu^2}{N_c^2L^2}\left(\sum_{\substack{k=0\\k\neq i}}^{N_C-1}\sum_{m=1}^{L}\sum_{n=0}^{N_C-1} b_{k,l}\left|\alpha_{i,l}\right|^2 e^{j2\pi\Delta f_k T\{n+\epsilon(n,m)\}}\right)^2 \ldots (17)$$

Variance of interference from other carriers of other users,

$$\sigma_{CUI}^2 = \frac{2P\mu^2}{N_c^2L^2}\left(\sum_{\substack{l=1\\l\neq l}}^{N_u}\sum_{m=1}^{L}\sum_{\substack{k=0\\k\neq i}}^{N_C-1}\sum_{n=0}^{N_C-1} b_{k,l}\left|\alpha_{i,l}\right|^2 e^{j2\pi\Delta f_k T\{n+\epsilon(n,m)\}}\right)^2 \ldots (18)$$

Variance of interference from same carrier of other users,

$$\sigma_{UI}^2 = \frac{2P\mu^2}{N_c^2L^2}\left(\sum_{m=1}^{L}\sum_{\substack{l=1\\l\neq l}}^{N_u}\sum_{n=0}^{N_C-1} b_{i,l}\left|\alpha_{i,l}\right|^2 e^{j2\pi\Delta f_k T\{n+\epsilon(n,m)\}}\right)^2 \ldots (19)$$

Variance of noise component is

$$\sigma_n^2 = \frac{N_0}{4T_b}\sum_{i=0}^{N_C-1}\left|\alpha_{i,l}\right|^2$$

and μ is the percentage of cross correlation among contributions from different components.

So, the SINR will be,

$$SINR = \frac{P_S\, sinc^2(\Delta f T_b)}{\sigma_{CI}^2 + \sigma_{CUI}^2 + \sigma_{UI}^2 + \sigma_n^2}$$

$$=> SINR = \frac{P_S\, sinc^2(\Delta f T_b)}{\sigma_{CI}^2 + \sigma_{CUI}^2 + \sigma_{UI}^2 + \frac{N_0}{4T_b}\sum_{i=0}^{N-1}\left|\alpha_{i,l}\right|^2} \ldots.(20)$$

For the Rayleigh fading channel when α is Rayleigh distributed ; pdf of α is :

$$p(\alpha) = \frac{\alpha}{\sigma_\alpha^2}\exp\left(-\frac{\alpha^2}{2\sigma_\alpha^2}\right) \ldots(21)$$

But, the pdf of zero mean Gaussian distributed jitter is,

$$p(\epsilon) = \frac{1}{\sqrt{2\pi\sigma_\epsilon^2}}\exp\left(-\frac{\epsilon^2}{2\sigma_\epsilon^2}\right) \ldots.(22)$$

Where, σ_ϵ^2 is the variance of the timing jitter.

The BER expression for BPSK is,

$$P_b(\epsilon,\alpha) = Q\left[\sqrt{SINR}\right] = \frac{1}{2}erfc\left(\sqrt{\frac{SINR}{2}}\right) \ldots..(23)$$

$$\text{Where,} \quad Q(x) = \frac{1}{\sqrt{2\pi}}\int_0^x e^{-\frac{y^2}{2}}dy$$

The unconditional or average BER can be calculated by averaging the conditional BER $P_b(\epsilon,\alpha)$, over all possible values of ϵ and α .

$$BER = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} P_b(\epsilon,\alpha)\, p(\alpha)\, p(\epsilon)\, d\alpha\, d\epsilon \ldots..(24)$$

For discrete case when number of subcarrier is limited,

$$BER = \sum_{i=0}^{N_C-1}\sum_{i=0}^{N_C-1} P_b(\epsilon_i,\alpha_i)p(\alpha_i)\, p(\epsilon_i) \ldots..(25)$$

IV. Simulation Results And Discussions

The BER performance results are evaluated using a bit rate of 1Mbps. For all of our computations, we keep some of the system parameters fixed e.g. fading variance=0.2, number of chips per bit, L=16 and cross-correlation of PN sequences=0.1

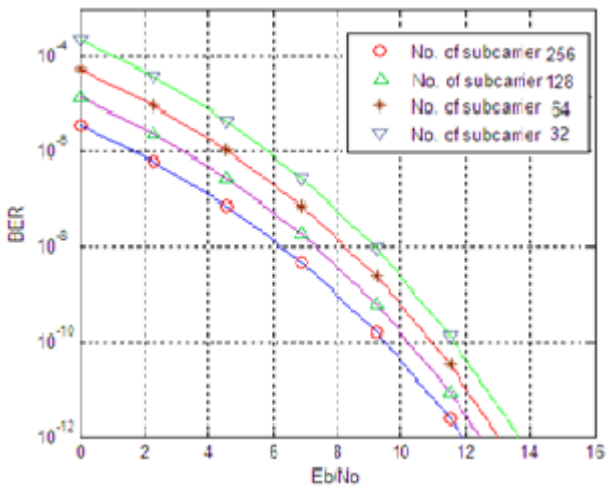


Fig. 3 BER as a function of Eb/No for different number of subcarriers

In Fig.3, BER is plotted against $Eb/N_0$with the number of subcarriers as a parameter. For this purpose, a range of subcarriers from 32 to 256 are used. Additionally, the number of users = 16 are kept fixed. From the figure we can see that, to achieve a BER of 10^{-8}, required Eb/N_0 is 9, 8, 1 7 and 6 dB respectively for 32, 64, 128 and 256 subcarriers. Increase in the number of subcarrier results reduction in transmission rate of symbols and noise bandwidth. This is always associated with an improvement of bit error performance. The figure indicates a tremendous impact of number of subcarriers on BER. It is observed that for higher values of E_b/N_0, there is greater improvement in BER with increase in number of subcarriers.

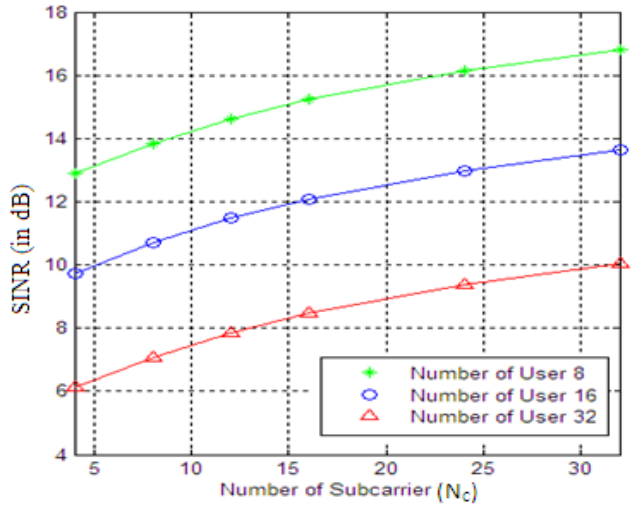


Fig. 4 SINR as a function of number of subcarriers for different number of users

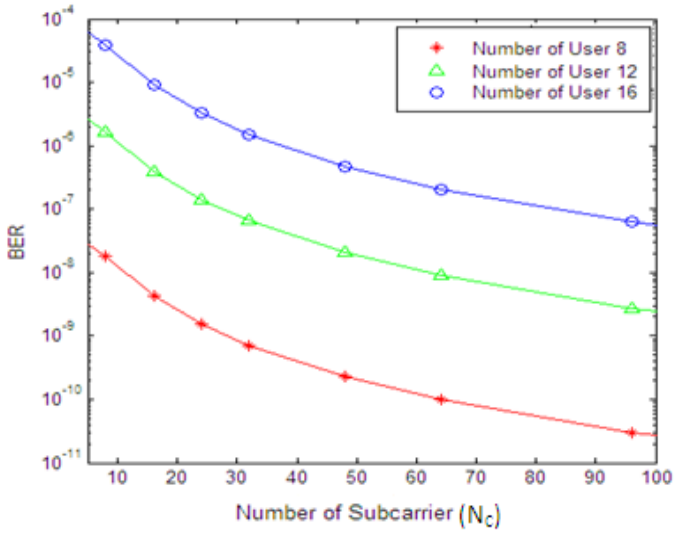


Fig. 5 BER as a function of number of subcarriers for different number of users

Again, Fig. 4, shows the SINR against number of subcarriers for different no. of users, N_u = 8, 16, 32. To achieve a certain SINR for definite number of users the minimum requirement of subcarriers is to be fulfilled. But, the maximum number of subcarrier is limited by the minimum inter-carrier separation requirement of the system which in turn limits the minimum achievable BER.

This relation for BER is reflected in Fig. 5 for different no. of users, N_u = 8, 12, 16.

Fig. 6 represents the plot of SINR vs. number of users, for a range of subcarriers and a given Eb/N_0. It shows that, SINR decreases significantly along with the number of users as we know the increased user means increased interference. Again, for a moderate number of users, SINR can be improved by using a number of subcarriers and in this plot the number of subcarrier is varied within a range of 8 to 64. An increment in the number of subcarriers improves SINR when there are a large number of users.

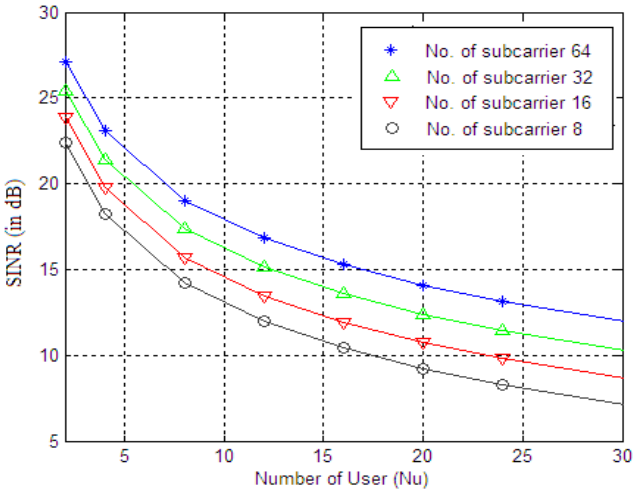


Fig. 6 SINR as a function of number of users for different number of subcarriers

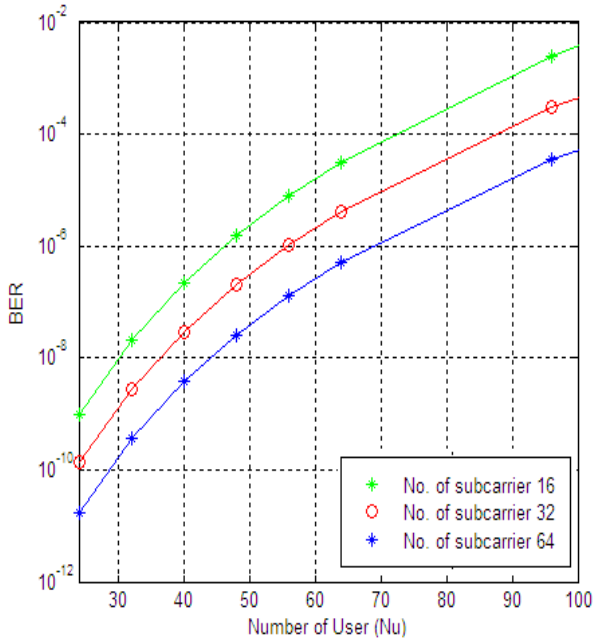


Fig. 7 BER as a function of number of users for different number of subcarriers

For the similar condition, Fig. 7 shows that BER increases substantially for the increment in user number which can be compensated by increasing number of subcarriers.

V. Conclusions

In this paper, we have investigated the effect of sampling timing jitters as well as of Carrier Frequency Offset on the BER performance of MC-DS-CDMA systems operating in multipath Rayleigh fading environment. Analytical expression for the MC-DS-CDMA decision variable in presence of timing jitter has been derived and various interference components caused by timing jitter have been clearly identified. We have also deduced expressions for SINR and BER for the system.

The number of subcarriers used is a very dominating factor for improving BER performance in the multicarrier system. For a better BER performance, the number of subcarrier should be increased and also reversely, increasing the subcarrier number requires less Eb/N_0 to achieve the desired BER. The number of users also affects BER performance very much.

Possible future research would be to extend the work of this paper to obtain analytical BER results considering the changes in jitter variance and random fluctuations in frequency offset. So that, more insights can be cast on the intrinsic mechanism of MC-DS-CDMA performance degradation.

Acknowledgement

We acknowledge the support of the department of Electrical & Electronic Engineering, Bangladesh University of Engineering & Technology to carry out this work as a part of final thesis of postgraduate program.

References

[1] S. Hara and R. Prasad, "Overview of multicarrier CDMA," *IEEE Commun Mag.*, vol. 35, no. 12, pp.126-133, Dec. 1997.

[2] X. Gui and T.S. Ng, "Performance of asynchronous orthogonal multicarrier CDMA system in frequency selective fading channel," *IEEE Trans. Commun.*, vol. 47, no. 7, pp.1084-1091, Jul. 1999

[3] L. Yang and L. Hanzo, "Multicarrier DS-CDMA: A multiple access scheme for ubiquitous broadband wireless communications," *IEEE Commun Mag.*, vol. 41, no. 10, pp.116–124, Oct. 2003.

[4] H. Steendam, "The effect of timing jitter on MC-DS-CDMA," *IEEE Trans. Commun.*, vol. 52, no. 3, pp. 467-472, Mar. 2004.

[5] X. Gui and L.Xu, "Performance of MC-DS-CDMA in the presence of timing jitter," in *Proc. WiMob'11*, pp.466-470, Wuhan, Oct. 2011.

[6] J.I.Z. Chen, "The Impact on Channel Correlation for MC-DS-CDMA System in Small-Scale Fading Environments," *Wireless Personal Commun Journal*, vol. 41, pp.471-485, 2006.

[7] Q. Chen, E. S. Sousa and S. Pasupathy, "Performance of Coded Multi-Carrier DS-CDMA System in Multi-Path Fading Channels," *Wireless Personal Commun. Journal*, vol. 2, pp 1-2, 1995.

[8] S.D. Paul, S.Asif and S. P. Majumder, "Performance Analysis of MC-DS-CDMA wireless communication system with Rake Receiver over a Rayleigh Fading Channel with Receive Diversity," in *Proc. ICECE'12*, pp.178-181, Dhaka, Dec. 2012.

[9] S. Mallick and S. P. Majumder, "Performance Analysis of an OFDM system in the presence of Carrier Frequency Offset, Phase Noise and Timing Jitter over Rayleigh Fading Channels," in *Proc. ICECE'08*, pp.205-210, Dhaka, Dec. 2008.

Intrusion Detection Learning Algorithm through Network Mining

A J M Abu Afza
Department of Computer Science and Engineering
Jahangirnagar University, Dhaka, Bangladesh
Email: afza22bd@gmail.com

Mohammad Shorif Uddin
Department of Computer Science and Engineering
Jahangirnagar University, Dhaka, Bangladesh
Email: shorifuddin@gmail.com

Abstract— This paper presents a learning algorithm for adaptive network intrusion detection based on clustering and naïve Bayesian classifier, which induces a hybridization of unsupervised and supervised learning processes. The proposed approach scales up the balance detection rates for different types of network intrusions, and keeps the false positives at acceptable level in network intrusion detection. The algorithm first clusters the network logs into several groups based on similarity of network logs, and then calculates the prior and class conditional probabilities for each cluster. In classifying a new network log, the algorithm calculates the similarity of attribute values of network data with each cluster and initialize a weight value for each cluster. Then each cluster classifies the network data with its priori and conditional probabilities that multiply with respective cluster's weight value. Finally, voting techniques applied for classifying the new network data based on each cluster's classification result. The performance of the proposed algorithm tested by employing KDD99 benchmark network intrusion detection dataset, and the experimental results proved that it improves the detection rates as well as reduces the false positives for different types of network intrusions.

Keywords- intrusion detection; boosting; naïve Bayesian classifier.

I. INTRODUCTION

Intrusion detection can be performed manually or automatically. Manual intrusion detection might take place by examining log files or other evidence for signs of intrusions, including network traffic. A system that performs automated intrusion detection is called an Intrusion Detection System (IDS). An IDS can be either host-based, if it monitors system calls or logs, or network-based if it monitors the flow of network packets. Modern IDSs are usually a combination of these two approaches. Another important distinction is between systems that identify patterns of traffic or application data presumed to be malicious (misuse detection systems), and systems that compare activities against a 'normal' baseline (anomaly detection systems). IDS collect information from a variety of network sources using intrusion detection sensors, and analyze the information for signs of intrusions that attempt to compromise the confidentiality and integrity of networks [1]-[3]. For this reason, Intrusion Detection Systems (IDS) have become important tools for ensuring network security. IDS were first introduced by James P. Anderson in 1980 [4]. In his report, Anderson presents a threat model that classifies intrusions to develop a security monitoring surveillance system based on detecting anomalies in user behavior. Later in 1986, Dorothy Denning [5], [6] proposed several models for IDS based on statistics, Markov chains, time-series, etc. IDS were first implemented for host-based that located in servers to examine the internal interfaces [7], but with the evolution of computer networks the focus gradually shifted toward network-based. Network-based intrusion detection systems (NIDS) monitor and analyze network traffics in the network for detecting intrusions from internal and external intruders [8]-[13]. Internal intruders are the inside users in the network with some authority, but try to gain extra ability to take action without legitimate authorization. External intruders are the outside users without any authorized access to the network that they attack. IDS notify network security administrator or automated intrusion prevention systems (IPS) about the network attacks, when an intruder try to break the network. Since the amount of audit data or network logs that an IDS needs to examine is very large even for a small network, several data mining algorithms, such as decision tree, naive Bayesian classifier, neural network, support vector machines, and fuzzy classification, etc [14]-[24] have been widely used by the IDS community for detecting known and unknown intrusions. Data mining based intrusion detection algorithms aim to solve the problems of analyzing the huge volumes of audit data and realizing performance optimization of detection rules [25]. But there are still some drawbacks in currently available commercial IDS, such as low detection accuracy, large number of false positives, unbalanced detection rates for different types of intrusions, long response time, and redundant input attributes. A conventional intrusion detection database is complex, dynamic, and composed of many different attributes. The problem is that not all attributes in intrusion detection database may be needed to build efficient and effective IDS. In fact, the use of redundant attributes may interfere with the correct completion of mining task, because the information they added is contained in other attributes. The use of all attributes may simply increase the overall complexity of detection model, increase computational time, and decrease the detection accuracy of the intrusion detection algorithms. It has been tested that effective attributes selection improves the detection rates for different types of network intrusions in intrusion detection. Intrusion detection is a process of gathering intrusion related knowledge that occurred in the computer networks or systems and analyzing them for detecting future intrusions. Intrusion detection can be broadly

978-1-4799-3497-3/13/$31.00 ©2013 IEEE

divided into two categories: misuse detection and anomaly detection. Misuse detection systems detect attacks based on known attack patterns that stored in a database, while anomaly detection systems detect deviations in activity from normal profiles. Misuse detection systems use various methods including rule-based expert systems, model-based reasoning systems, state transition analysis, genetic algorithms, fuzzy logic, and keystroke monitoring. The main advantage of misuse detection system is to produce very low false positives (FP), but it requires regular updates of rules, and not capable of detecting unknown or new intrusions. On the other hand, anomaly detection systems detect deviations from normal behavior, and based on a threshold value determines if it is normal behavior or intrusion [26]. It can detect known and unknown or new intrusions. There are various approaches for anomaly detection including statistical analysis, neural networks, machine learning, and artificial immune system. The main disadvantage of anomaly detection is that it provides many false positives (FP). The success of an IDS can be characterized in both detection rates (DR) and false positives (FP) for different types of intrusions [27]. Ideally, IDS should have an attack detection rate of 100% along with false positive of 0%, which is really hard to achieve. To apply data mining techniques in intrusion detection, first the collected network logs or audit data needs to be preprocessed and converted to the format that suitable for mining. Next, the reformatted data will be used to develop a clustering or classification model. Data mining provide decision support for intrusion management, and also help IDS for detecting new vulnerabilities and intrusions by discovering unknown patterns of attacks or intrusions.

II. CLASSIFIER CONSTRUCTION TECHNIQUES

A. Naïve Bayesian Classifier

Naïve Bayesian (NB) classifier is a simple probabilistic classifier based on probability models that incorporate strong independence assumptions which often have no bearing in reality. The probability model can be derived using Bayes rule. Depending on the precise nature of the probability model, NB classifier can be trained very efficiently in a supervised learning. In many practical applications, parameter estimation for naïve Bayesian models uses the method of maximum likelihood. In spite of naïve design and apparently oversimplified assumptions, naïve Bayesian classifiers often work much better in many complex real-world situations. The NB classifier is given as input a set of training examples each of which is described by attributes A_1 through A_K and an associated class, C. The objective is to classify an unseen example whose class value is unknown but values for attributes A_1 through A_K are known and they are $a_1, a_2, \ldots, a_K$ respectively. The optimal prediction of the unseen example is the class value c such that $P(C=c_i|A_1 = a_1, \ldots A_K = a_K)$ is maximum. By Bayes rule this probability equals to

$$\arg\max = \frac{P(A_1 = a_1 \ldots\ldots A_k = a_k \mid C = c_1)}{P(A_1 = a_1 \ldots\ldots A_k = a_k)} P(\mid C = c_1) \quad (7)$$

Where $P(C=c_i)$ is the prior probability of class c_i, $P(A_1= a_1, \ldots A_K = a_K)$ is the probability of occurrence of the description of a particular example, and $P(A_1= a_1, \ldots A_K = a_K|C= c_i)$ is the class conditional probability of the description of a particular example c_i of class C. The prior probability of a class can be estimated from training data. The probability of occurrence of the description of particular examples is irrelevant for decision making since it is the same for each class value c. Learning is therefore reduced to the problem of estimating the class conditional probability of all possible description of examples from training data. The class conditional probability can be written in expanded from as follows:

$$\begin{aligned} &P(A_1 = a_1 \ldots\ldots A_k = a_k \mid C = c_k) \\ &\quad = P(A_1 = a_1 \mid A_2 = a_2 \lambda \ldots\ldots A_k = a_k C = c_i) \\ &\quad * P(A_2 = a_2 \mid A_3 = a_3 \lambda \ldots\ldots A_k = a_k C = c_i) \\ &\quad * P(A_3 = a_3 \mid A_4 = a_4 \lambda \ldots\ldots A_k = a_k C = c_i) \\ &\quad * P(A_4 = a_4 \lambda \ldots\ldots A_k = a_k C = c_i) \end{aligned} \quad (8)$$

In NB, it is assumed that outcome of attribute A_1 is independent of the outcome of all other attributes A_j, given c. Thus class conditional probabilities become:

$$P(A_1 = a_1 \ldots A_k \setminus C = c_i) = \prod_{i=1}^{k} P(A_i = a_i \ldots A_k \setminus C = c_i) \quad (9)$$

If the above value is inserted in equation "(7)" it becomes:

$$\equiv \arg\max \in CP(C = c) \prod_{i=1}^{k} P(A_i = a_i \ldots A_k \setminus C = c_i) \quad (10)$$

In Naïve Bayesian learning, the probability values of equation "(10)" are estimated from the given training data. These estimated values are then used to classify unknown examples.

B. K-Nearest Neighbors

The K nearest neighbors (KNN) is a classification algorithm based on the use of distance measures. It finds k examples in training data that are closest to the test example and assigns the most frequent label among these examples to the new example. When a classification is to be made for a new example, its distance to each attribute in the training data must be determined. Only the K closest examples in the training data are considered further. The new example is then placed in the class that contains the most examples from this set of K closest examples. KNN can be considered decision making technique as equivalent to Bayesian classifier in which the number of neighbors of each example is used as an estimate of the relative posterior probabilities of class membership in the neighborhood of a sample to be classified.

978-1-4799-3497-3/13/$31.00 ©2013 IEEE

C. Clustering

Clustering is an unsupervised learning method, where unlabeled data is used to train the model. The main idea is that the clustering method groups the whole instances of dataset into subsets (clusters) of instance, based on the similarity (or dissimilarity) of instances. There are many measures to estimate the distance between any pair of instances (data points). Some common distance measures found in leterature are Minkowski, Euclidean, Chebychev, Cosine, Pearson Correlation, Jaccard, and Dice Coeeficient. Clustering method is widely used in many domains. However, the accuracy of clustering is dependent on algorithms used to cluster data points. Several algorithms of clustering are proposed in literature. Some common clustering algorithms are known as hierarchical clustering (i.e., sigle-link, complete-link, and average-link clustering), k-means clustering, graph-theoretic clustering, density-based clustering, SOM neural network, fuzzy clustering, and so on.

III. PROPOSED ALGORITHM

Here we present a new decision tree based learning algorithm, ACDT (Attacks Classification using Decision Tree) algorithm to address the problem of attacks classification in intrusion detection for classifying different types of network attacks. ACDT algorithm improves the 98% detection rates (DR) and reduces false positives (FP) using KDD99 benchmark network intrusion detection dataset in comparison with other existing methods.

Following describes the proposed algorithms:

Algorithm

Input: Database, D

Output: Intrusion Detection Model

Learning Algorithm:

Step 1. for each example t_i D, check the similarity of examples: $sim(t_i, t_l)$;

Step 2. Put examples into cluster: $D_i \leftarrow t_i$;

Step 3. for each cluster D_i, calculate the prior probabilities: $P(C_j) = \dfrac{\sum t_{i \to C_j}}{\sum_{i=1}^{n} t_i}$;

Step 4. for each cluster D_i, calculate the conditional probabilities:

$$P(A_{ij}|C_j) = \frac{\sum_{i=1}^{n} A_{i \to C_j}}{\sum t_{i \to C_j}};$$

Step 5. for each cluster D_i, store the prior probabilities, $S_1 = P(C_j)$; and conditional probabilities, $S_2 = P(A_{ij}|C_j)$;

Step 6. For classifying new example, find the similarity of attribute values of classified example with each cluster and initialized the weight value for each cluster in D, w_i=1/n, where n is the total number of the examples in a cluster.

Step 7. Classify the new example with their prior and conditional probabilities that multiply the respective cluster's weight value.

Step 8. Using voting technique of each cluster result, classify the new example.

IV. EXPERIMENTAL RESULT

The KDD 1999 cup benchmark intrusion detection dataset was used in the 3rd International Knowledge Discovery and Data Mining Tools Competition for building a network intrusion detector, a predictive model capable of distinguishing between intrusions and normal connections [35]. In 1998, DARPA intrusion detection evaluation program, a simulated environment was set up to acquire raw TCP/IP dump data for a local-area network (LAN) by the MIT Lincoln Lab to compare the performance of various intrusion detection methods. It was operated like a real environment, but being blasted with multiple intrusion attacks and received much attention in the research community of adaptive intrusion detection. The KDD99 dataset contest uses a version of DARPA98 dataset. In KDD99 dataset, each example represents attribute values of a class in the network data row, and each class is labeled either normal or attack. The classes in KDD99 dataset can be categorized into five main classes (one normal class and four main intrusion classes: probe, DOS, U2R, and R2L).

1. Normal connections are generated by simulated daily user behavior such as downloading files, visiting web pages.
2. Denial of Service (DoS) attack causes the computing power or memory of a victim machine too busy or too full to handle legitimate requests. DoS attacks are classified based on the services that an attacker renders unavailable to legitimate users like apache2, land, mail bomb, back, etc.
3. Remote to User (R2L) is an attack that a remote user gains access of a local user or account by sending packets to a machine over a network communication, which include send-mail, and Xlock.
4. User to Root (U2R) is an attack that an intruder begins with the access of a normal user account and then becomes a root-user by exploiting various vulnerabilities of the system. Most common exploits of U2R attacks are regular buffer-overflows, load-module, Fd-format, and Ffb-config.
5. Probing (Probe) is an attack that scans a network to gather information or find known vulnerabilities. An intruder with a map of machines and services that are available on a network can use the information to look for exploits.

In KDD99 dataset these four attacks (DoS, U2R, R2L, and probe) are divided into 22 different attacks that tabulated in table 1.

978-1-4799-3497-3/13/$31.00 ©2013 IEEE

Table 1 Different type of attacks in KDD99 Dataset.

4 Main Attack Classes	Type of Attributes
Denial of Service (DoS)	back, land, neptune, pod, smurt, teardrop
Remote to User (R2L)	ftp write, guess passwd, imap, multihop, phf, spy,warezclient, warezmaster
User to Root (U2R)	buffer overflow, perl, loadmodule, rootkit
Probing	ipsweep, nmap, portsweep, satan

There are total 41 input attributes in KDD99 dataset for each network connection that have either discrete or continuous values and divided into three groups. The first group of attributes is the basic features of network connection, which include the duration, prototype, service, number of bytes from source IP addresses or from destination IP addresses, and some ags in TCP connections. The second group of attributes in KDD99 is composed of the content features of network connections and the third group is composed of the statistical features that are computed either by a time window or a window of certain kind of connections. Table 2 show the number of examples in 10% training and testing data of KDD99 dataset.

Table 2 Number of training and testing examples in KDD99 Dataset.

Attack Types	Training Examples	Testing Examples
Normal	9727	60592
Denial of Service	39145	237594
Remote to User	1126	8606
User to Root	52	70
Probing	4107	4166
Total Examples	494020	311028

In the experiments, first we performed the classification on KDD99 testing data using Naïve Bayesian classifier (NB) and Intrusion Detection Learning Algorithm and compare their results shown in Table 3. It is clear that the proposed new algorithm is approximately 2 times faster in training and testing than conventional algorithms.

Table 3 Testing and Training Time Comparison

Algorithm	Training Time (s)	Testing Time (s)
Naïve Bayesian Classifier (NB)	42.9	18.4
Proposed Algorithm	24.6	7.1

The performance comparison between NB and Proposed Algorithm on KDD99 (using 41 attributes) testing data is listed in Table 4, which shows that the Proposed Algorithm performs balanced and high classification rates on 5 attack classes of KDD99 data and minimize the false positives.

Table 4 Performance Comparison between NB and Proposed Algorithm

	Normal	Probe	DoS	U2R	R2L
Naïve Bayesian Classifier (DR %)	99.25	99.13	99.69	64	99.11
Naïve Bayesian Classifier (FP %)	0.08	0.45	0.04	0.14	8.02
Proposed Algorithm (DR %)	99.42	99.20	99.59	99.10	99.14
Proposed Algorithm (FP %)	0.03	0.23	0.03	0.08	6.21

We also compare the detection performance among Support Vector Machines (SVM), Neural Network (NN), Genetic Algorithm (GA), Naïve Bayesian Classifier (NB), and improved self adaptive Bayesian algorithm (ISABA) on KDD99 dataset. In total, 41 attributes of KDD99 dataset have been used. Each connection can be categorized into five main classes (one normal class and four main intrusion classes: probe, DoS, U2R, and R2L). The experimental setting uses 494020 data samples for training and 311028 data samples for testing. The comparative results are summarized in Table 5.

Table 5 Comparison of several algorithms

Classes	SVM	NN	GA	NB	Proposed Algorithm
Normal	99.4	99.6	99.3	99.55	99.85
Probe	89.2	92.7	98.46	99.43	99.75
DoS	94.7	97.5	99.57	99.69	99.61
U2R	71.4	48	99.22	64	99.51
R2L	87.2	98	98.54	99.11	99.45

V. CONCLUSION

In this paper, we present a Dependable Network Intrusion Detection System (DNIDS) by integrating detection method with an intrusion-tolerant mechanism. To address the detection issue, we propose a Combined Strangeness and Isolation measure K-Nearest Neighbor (CSI-KNN) algorithm. The algorithm employs a combined model that uses two different measures to improve its detection ability. The strangeness measure is good at detecting attacks that show more similarity to known attacks than to normal behaviors. The isolation measure is designed to detect those unknown attacks that behave significantly differently from normal activities. By combining the two measures, our CSI-KNN algorithm possesses the ability to detect more attacks to detect more attacks than either alone. Moreover, the algorithm can provide a graded confidence that indicates the reliability of the classification. We also propose an NIDS framework that employs multiple independent service-based CSI-KNN classifiers. The framework possesses exibility since the classifiers can be distributed across different hosts. We implemented the CIS-KNN algorithm and tested the

978-1-4799-3497-3/13/$31.00 ©2013 IEEE

detection approach over the KDD'99 benchmark dataset. The experimental results show that the performance of our approach is better than the best result from KDD'99 contest. We have also compared our results with one of the most recent approaches that use the KDD'99 dataset. The future research issue will be applying this classifier in classification problems of real world problem domains.

ACKNOWLEDGMENT

Support for this research received from Department of Computer Science and Engineering, Jahangirnagar University Savar, Dhaka, Bangladesh.

REFERENCES

[1] X. D. Hoang, J. Hu, and P. Bertok, "A program-based anomaly intrusion detection scheme using multiple detection engines and fuzzy inference", *Journal of Network and Computer Applications*, vol. 32, no. 6, pp. 1219-1228, November 2009.

[2] P. G. Teodoro, J. D. Verdijo, G. M. Fernandez, and E. Vazquez, "Anomaly-based network intrusion detection: Techniques, systems and challenges", *Computers & Security*, vol. 28, no. 1-2, pp. 18-28, February - March 2009.

[3] A. Patch and J. M. Park, "An overview of anomaly detection techniques: Existing solutions and latest technological trends", *Computer Netwroks*, vol. 51, no. 12, pp. 3448-3470, August 2007.

[4] J. P. Anderson, "Computer security threat monitoring and surveillance", *Technical report 98-17, James P. Anderson Co., Fort Washington, Pennsylvania*, USA, April 1980.

[5] D. E. Denning, "An intrusion detection model", *In Proc. of the Seventh IEEE Symposium on Security and Privacy*, pp. 119-131, May 1986.

[6] D.-E. Denning, "An intrusion detection model", *IEEE Transaction on Software Engineering*, vol. 13, no. 2, pp. 222-232, 1987.

[7] D. Y. Yeung and Y. X. Ding, "Host-based intrusion detection using dynamic and static behavioral models", *Pattern Recognition*, vol. 36, pp. 229-243, 2003.

[8] L. C.Wuu, C. H. Hung, and S. F. Chen, "Building intrusion pattern miner for snort network intrusion detection system", *Journal of Systems and Software*, vol. 80, no. 10, pp. 1699-1715, October 2007.

[9] C. M. Chen, Y. L. Chen, and H. C. Lin, "An efficient network intrusion detection", *Computer Communications*, vol. 33, no. 4, 1 March 2010.

[10] M. A. Aydin, A. H. Zaim, and K. G. Ceylan, "A hybrid intrusion detection system design for computer network security", *Computers & Electrical Engineering*, vol. 35, no. 3, pp. 517-526, May 2009.

[11] F. Seredynski and P. Bouvry, "Anomaly detection in tcp/ip networks using immune systems paradigm", *Computer Communications*, vol. 30, no. 4, pp. 740-749, 26 February 2007.

[12] J. J. C. Foster, M. Jonkman, R. Marty, and E. Seagren, "Intrusion detection systems", *Snort Intrusion detection and Prevention Toolkit*, pp. 1-30, 2006.

[13] B. Rexworthy, "Intrusion detections systems an outmoded network protection model", *Network Security*, vol. 2009, no. 6, pp. 17-19, June 2009.

[14] W. Wang, X. Guan, and X. Zhang, "Processing of massive audit data streams for real-time anomaly intrusion detection", *Computer Communications*, vol. 31, no. 1, pp. 58-72, 15 January 2008.

[15] H. C. Wu and S. H. S. Huand, "Neural network-based detection of stepping-stone intrusion", *Expert Systems with Applications*, vol. 37, no. 2, pp. 1431-1437, March 2010.

[16] X. Tong, Z. Wang, and H. Yu, "A research using hybrid rbf/elman neural networks for intrusion detection system secure model", *Computer Physics Communications*, vol. 180, no. 10, pp. 1795-1801, October 2009.

[17] C. Forn and C. Y. Lin, "A triangle area based nearset neighbors approach to intrusion detection", *Pattern Recognition*, vol. 43, no. 1, pp. 222-229, January 2010.

[18] K. Sha_ and H. A. Abbass, "An adaptive genetic-based signature learning system for intrusion detection", *Expert Systems with Applications*, vol. 36, no. 10, pp. 12 036-12 043, December 2009.

[19] Z. Bankovic, D. Stepanovic, S. Bojanic, and O. N. Talasriz, "Improving network security using genetic algorithm approach", *Computers & Electrical Engineering*, vol. 33, no. 5-6, pp. 438-541, 2007.

[20] Y. Li and L. Guo, "An active learning based tcm-knn algorithm for supervised network intrusion detection", *Computer & Security*, vol. 26, no. 7-8, pp. 459-467, December 2007.

[21] W. H. Chen, S. H. Hsu, and H. P. Shen, "Application of svm and ann for intrusion detection", *Computers & Operations Research*, vol. 32, no. 10, October 2005.

[22] M. Y. Su, G. J. Yu, and C. Y. Lin, "A real-time network intrusion detection system for large-scale attacks based on an incremental mining approach", *Computers & Security*, vol. 28, no. 5, pp. 301-309, July 2009.

[23] Z. Jinquan, L. Xiaojie, L. Tao, L. Caiming, P. Lingxi, and S. Feixian, "A selfadaptive negative selection algorithm used for anomaly detection", *Progress in Natural Science*, vol. 19, no. 2, pp. 261-266, February 2009.

[24] Z. Zhang and H. Shen, "Application of online-training svms for real-time intrusion detection with different considerations", *Computer Communications*, vol. 28, no. 12, pp. 1428-1442, July 2005.

[25] S. Y. Wu and E. Yen, "Data mining based intrusion detection", *Expert System with Applications*, vol. 36, no. 3, pp. 5605-5612, April 2009.

[26] A. Lazarevic, L. Ertoz, V. Kumar, A. Ozgur, and J. Srivastava, "A comparative study of anomaly detection schemes in network intrusion detection", *In Proc. of the SIAM Conference on Data Mining*, 2003.

[27] M. Stillerman, C. Marceau, and M. Stillman, "Intrusion detection on distributed systems", *Communications of the ACM*, vol. 42, pp. 62-69, 2007.

[28] P. A. Poras and A. Valdes, "Live tra_c analysis of tcp/ip gateways", *In Proc. of the Network and Distributed System*

Security Symposium, San Diego, CA: Internet Society, 11-13 March 1998.

[29] T. D. Garvey and T. F. Lunt, "Model based intrusion detection", *In Proc. of the 14th National Conference Security Conference*, pp. 372-385, 1991.

[30] F. Carrettoni, S. Castano, G. Martella, and P. Samarati, "Retiss: A real time security system for threat detection using fuzzy logic", *In Proc. of the 25th IEEE International Carnahan Conference on Security Technology, Taipei, Taiwai ROC*, 1991.

[31] T. F. Lunt, A. Tamaru, F. Gilham, R. Jagannathan, P. G. Neumann, H. S. Javitz, A. Valdes, and T. D. Garvey, "A real-time intrusion detection expert system (ides)", *technical Report, Computer Science Laboratory, Menlo Park, CA: SRI International*.

[32] S. A. Hofmeyr, S. Forrest, and A. Somayaji, "Intrusion detection using sequences of system calls", *Journal of Computer Security*, vol. 6, pp. 151-180, 1998.

[33] S. A. Hofmeyr and S. Forrest, "Immunity by design: An arti_cial immune system", *In Proc. of the Genetic and Evolutionary Computation Conference (GECCO 99), San Mateo, CA: Morgan Kaufmann*, vol. 2, pp. 1289-1296, 1999.

[34] J. M. J. Bonifacio, A. M. Cansian, A. C. P. L. F. Carvalho, and E. S. Moreira, "Neural networks applied in intrusion detection systems", *In the Proc. of the International Conference on Computational Intelligence and Multimedia Application, Gold Coast, Australia*, pp. 276-280, 1997.

[35] "The kdd archive. kdd99 cup dataset", 1999, http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html. 24

978-1-4799-3497-3/13/$31.00 ©2013 IEEE

Two Stage Fuzzy Logic Based Clustering Approach Wireless Sensor Network LEACH Protocol

Md. Mustafa Kamal, Shamim Ara Shawkat and Mohammad Shah Alam
Institute of Information & Communication Technology (IICT)
Bangladesh University of Engineering and Technology (BUET)
Dhaka, Bangladesh
mustafacomp@gmail.com, msara@iict.buet.ac.bd and smalam@iict.buet.ac.bd

***Abstract*—Energy proficiency is an essential feature to intend an extensive existence wireless sensor network. More than a few methods have been proposed to boost the battery lifetime of sensor nodes. One of the methods is clustering approach which consists of two ways: selecting cluster heads with more residual energy and rotating cluster heads periodically from cluster to cluster thus extends the network lifetime. In this paper we present the wireless sensor networks by means of TSFL-LEACH (Two Stage Fuzzy Logic based Low Energy Adaptive Clustering Hierarchy) where completely two separate cluster levels are defined depending on the distance from the cluster heads and the cluster nodes. We also applied fuzzy asymmetrical clustering mechanism to prolong the lifetime of WSNs. We compare our algorithm with some popular algorithms namely LEACH, LEACH-C and EEUC. The simulation result shows that our bespoken edition improves the energy efficiency and enhances the life length of a wireless sensor network (WSN).**

Keywords*— *Energy efficiency, two stage clustering, residual energy, TSFL-LEACH, rotating cluster heads.

I. Introduction

There have been modern advances in micro-electromechanical systems (MEMS) technology, wireless communications, and digital electronics. These move forward have enabled the growth of low-cost, low-power, multifunctional sensor nodes that are small in size and communicate with each other using radio frequencies [1]. A single sensor node has limited capability in sensing and is not sufficient for gathering useful information from a specific domain. This data assembly process can be talented by the cooperative work of a number of sensor nodes. These collaboratively working sensor nodes form a network which is called a wireless sensor network (WSN).

The low costs make it potential to have a network of hundreds or thousands of sensors, enhancing the reliability and correctness of data [2]. A big improvement of sensor networks is represented by simplicity of exploitation, dropping installation cost, possibility to distribute the tiny sensors over a wide region, a larger fault tolerance because the failure of a single node not affects the network operation. Another important characteristic of sensor networks is represented from the capability to self configuring. There are a few of limitations of wireless sensor networks such as low bandwidth, limited resources as energy, etc. Usually these sensor networks are deployed in places where it is very difficult replace the individual nodes and for this reason it is desirable to realize communication protocols that increase the longevity of the network and it is preferable that nodes die together to replace all the nodes (the whole network).

The data being sensed by the nodes in the network must be transmitted to a control center or base station, where the end-user can access the data. There are many possible models for these micro sensor networks. In this work, we consider micro sensor networks where the base station is fixed and located far from the sensors. All nodes in the network are homogeneous and energy constrained. Thus, communication between the sensor nodes and the base station is expensive, and there are no "high-energy" nodes through which communication can proceed. These networks should function for as long as possible. It may be inconvenient or impossible to recharge node batteries. Therefore, all aspects of the node, from the hardware to the protocols, must be designed to be extremely energy efficient.

We use a simplified model shown in [4] for the radio hardware energy dissipation. Both the free space (d^2 power (loss) and the multi-path fading (d^4 power loss) channel models are used in the model, depending on the distance between the transmitter and receiver.

The energy spent for transmission of a l-bit packet over distance d is:

$$E_{Tx}(l,d) = \begin{cases} lE_{elec}^{Tx} + l\varepsilon_{fs}d^2; & d < d_0 \\ lE_{elec}^{Tx} + l\varepsilon_{mp}d^4; & d \geq d_0 \end{cases} \quad (1)$$

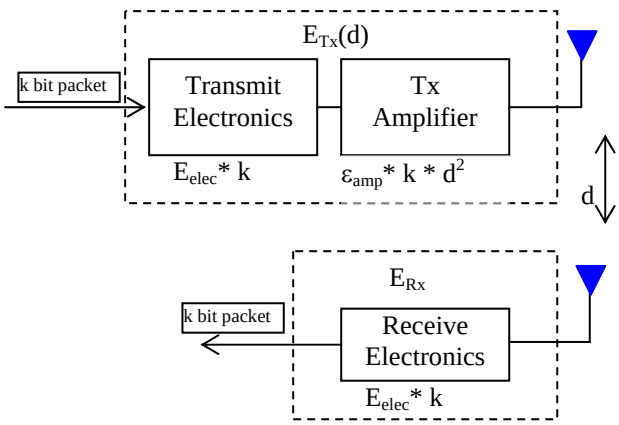


Fig 1. First order radio model.

The threshold value d_0 could be obtained via

$$d0 = \sqrt{\frac{\varepsilon_{fs}}{\varepsilon_{mp}}} \qquad (2)$$

ε_{fs} *and* ε_{mp} are the energy consumption factor of amplification for the free space and multi path ratio models, respectively. To receive this message, the radio expends energy:

$$E_{Rx}(l) = lE_{elec}^{Rx} \qquad (3)$$

A sensor node also consumes EDA (nJ/bit/signal) amount of energy for data aggregation. It's also assumed that the sensed information is highly correlated, thus the cluster head can always aggregate the data gathered from its members into a single length-fixed packet.

LEACH [4] is a probability based clustering protocol for periodical data gathering applications in WSN. In LEACH the whole operation is divided into rounds. Each round consists of a set up phase followed by steady state phase. During the set up phase, nodes decide on their own to be a CH or not by evaluating a certain probability without negotiating with other nodes and then broadcast an advertise message to the rest of the nodes. The non CH node examines the received advertise signal strength and decide of which CH node it will be cluster member for that round. Any node can be selected as the CH once in a 1/ P rounds, where P is the desired percentage of the CH in any round. In steady state phase CH aggregates data from the cluster members and sends the composite data to the base station by single hope communication[5]. Since in each round a new set of CHs are chosen randomly, LEACH evenly distributes the energy consumption of the CHs. Among the class of distributed, dynamic and randomized protocol, LEACH has the highest dominance due to its effectiveness in providing load balancing, scalability and energy efficient solution [6]. However it has some limitation. Authors at have shown that due to the random CH selecting strategy, the number of CH resulted by LEACH and other protocol in its class is not guaranteed to be equal to the expected optimal value and the probability that there is only one CH or there is no CH is high when the desired value of CH is small.

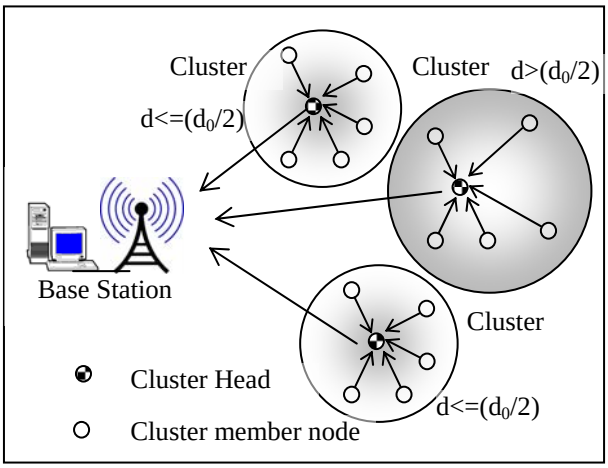


Fig 2. Cluster based system model of the proposed protocol.

So when no CH is elected, clustering structure is broken; every node must communicate with the BS directly[7]. Moreover, when few numbers, for example only one CH is elected, its energy will be drained rapidly. Therefore, the variability in the number of CHs adversely affects the energy efficiency, fairness among nodes and the system life time. Also it uses cluster formation in every round which increases the overhead of the protocol[8]. Lastly in the CH selection procedure any node select itself as the CH does not ensure that it's residual energy sustain for the whole round.

In section II clustering issues in sensor networks are considered. In section III is presented the fuzzy logic of TSFL-LEACH. Section IV discusses the details of our approach and the difference between this and the original LEACH. Section V presents simulation results and in section VI conclusions are done.

II. Proposed TSFL-LEACH Clustering Approach

This section provides a summary of the two-stage clustering technique relevant to the present work. In this section, we describe our proposed clustering algorithm TSFL-LEACH. This clustering technique is a distributed competitive unequal clustering algorithm similar to EEUC (Energy Efficient Unequal Clustering) [2]. It makes local decisions to establish radius and to elect cluster-heads. The main distinction between EEUC and TSFL-LEACH is their computation radius assessment technique. EEUC only considers distance from cluster to the base station to determine the radius. TSFL-LEACH exploits the threshold value (d_0). Moreover, TSFL-LEACH takes benefit of using fuzzy logic to calculate computation radius. CHEF (Cluster Head Election mechanism using Fuzzy logic) is a fuzzy approach for electing tentative cluster-head. LEACH protocol rotates the cluster heads periodically in each round by using a probabilistic model. Radius computation used in our proposed TSFL-LEACH is accomplished by using predefined fuzzy if-then mapping rules to handle the uncertainty. After forming the clusters we separate the clusters into two completely isolated stages one is called 1st stage and the other is called 2nd stage depending on the threshold value which is shown in the Figure 3. The radius of all the clusters in the first stage less than or equal to the half of the threshold value(d_0)

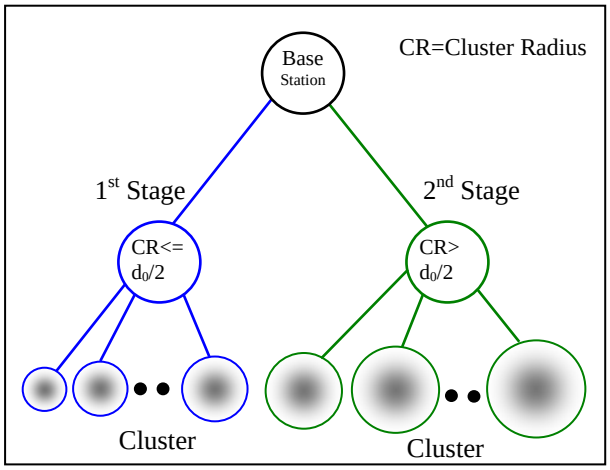


Fig 3. Overlapped sensor M within the cluster 1 and 2.

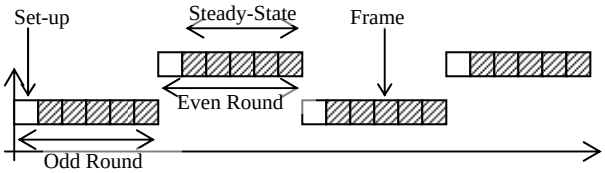


Fig 4. Time-line showing TSFL-LEACH protocol. Cluster are formed during the set-up phase and data transfers occur during the steady state phase.

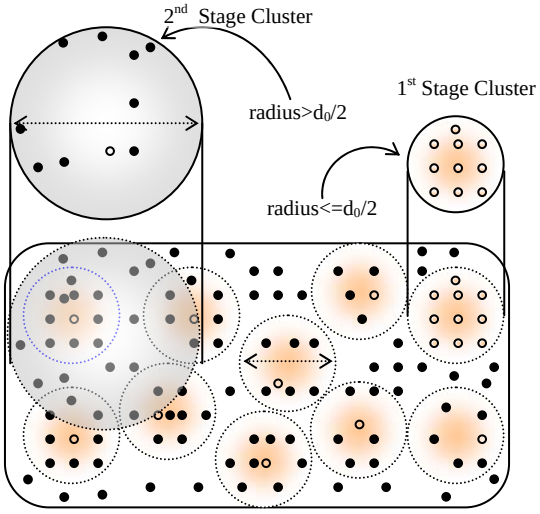


Fig 5. Two stage clustering for randomly placed sensor.

which means that the message transfer between the cluster head and sensor node requires only free space model(d^2 power loss). It does not require any multi path fading channel model (d^4 power loss). As a result we can save a lot of sensor energy because 1st stage clusters use only d^2 power loss. On the other hand, radius of the clusters in the second stages greater than half of the threshold value d_0 and the 2nd stage clusters work like LEACH protocol which means both d^2 and d^4 power loss model work together depending on the selection of the cluster head.

Figure-4 shows the time line of TSFL-LEACH protocol. In LEACH protocol the set-up phase and steady-state phase happened sequentially one after another but in our proposed method we divide this sequence into two different sequences one called odd round that is employed for 1st stage clusters and another called even round which is employed for 2nd stage clusters. In the odd round the 1st stage clusters send information to the base station and in the even round the 2nd stage clusters send information to the base station. When odd round clusters work, 2nd stage clusters save its energy and during the even round 1st stage clusters save its energy.

Figure 5 describes the detail diagram of two stage clustering. Here we see that the 1st stage cluster radius is less than or equal to ($d_0/2$) and 2nd stage cluster radius is greater than or equal to ($d_0/2$).

III. Proposed TSFL-LEACH Fuzzy Logic

In TSFL-LEACH cluster-head computation radius calculation, we use two fuzzy input variables. The first one is the distance from a particular tentative cluster-head to the base station. The fuzzy set that describes the distance to the base station is depicted in Figure 6. The linguistic variables for this fuzzy set are close, medium and far. We choose a trapezoidal membership function for close and far. On the other hand, the membership function of medium is a triangular membership function.

The second fuzzy input variable is cluster head energy of the tentative cluster-head. The fuzzy set that describes cluster head energy is illustrated in Figure 7. Low, Medium and High are the linguistic variables of this fuzzy set. Low and High linguistic variables have a trapezoidal membership function while Medium has a triangular membership function.

The only fuzzy output variable is the computation radius of the tentative cluster-head. Fuzzy set for computation radius fuzzy output variable is demonstrated in Figure 8. We have 9 linguistic variables which are Very Small, Small, Rather Small, Medium Small, Medium, Medium Large, Rather Large, Large and Very Large. Very Small and Very Large have a trapezoidal membership function. The remaining linguistic variables are represented by using triangular membership functions. The function in Figure 3 is not a symmetric triangular function as in Figure 6 and Figure 7. This is because the function in Figure 8 provides better results in our experiment. If a particular tentative cluster-head's battery is full and is located at the maximum distance to the base station, then it has the maximum computation radius. On the contrary, if a particular cluster-head's battery is near empty and is the closest distance to the base station, then it has the minimum computation radius. The remaining intermediate possibilities fall between these two extreme cases. The maximum computation radius is a static parameter for a particular WSN.

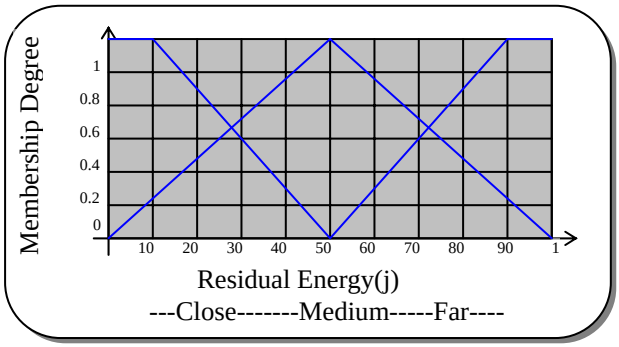


Fig 6. Fuzzy set for fuzzy input variable distance to base.

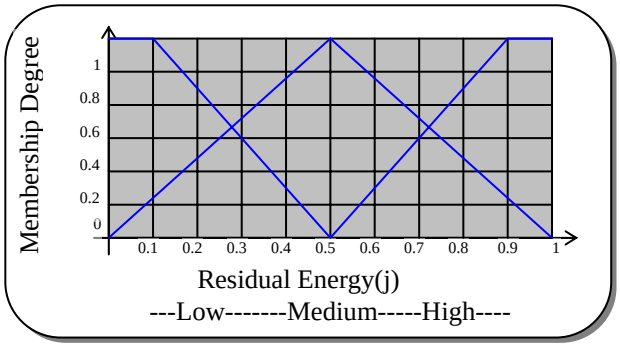


Fig. 7. Fuzzy set for fuzzy input variable Cluster Head Energy.

The base station broadcasts the value of this parameter to the entire network. Thus, all the sensor nodes know the maximum computation radius in advance. Each of the sensor nodes can calculate their relative computation radius according to the value of this parameter. The maximum distance to the base station is also a static parameter, because we assume that the sensor nodes are stationary. Each sensor node can determine their relative position to the base station considering the maximum distance to the base station in the WSN. After each tentative cluster-head determines its computation radius, cluster-head computation begins.

Each tentative cluster-head advertises Candidate CH Message to compute other tentative cluster-heads locally. This message is advertised to the tentative cluster-heads which are inside the maximum cluster-head computation radius. It includes node ID, computation radius and cluster head energy level of the source node. If a tentative cluster-head receives a Candidate CH Message from another tentative cluster-head which is in its computation range and the remaining energy of the source node is greater than the remaining energy of the receiving node, then the receiving node quits cluster-head computation and broadcasts a Quit Election Message.

If a particular tentative cluster-head has the highest remaining energy level among the tentative cluster-heads which it receives a Candidate CH Message from, then it becomes a cluster head. This computation guarantees that there does not exist another cluster-head in the computation radius of a particular cluster-head. After cluster-head election is completed, each ordinary sensor node joins to the closest cluster like in LEACH, CHEF and EEUC. Figure 4 illustrates a WSN which is clustered by using EAUCF algorithm. In this example, the number of deployed sensors is 200, Network size 200 x 200, base station location (100, 270), Initial energy 0.5 J, Data packet size 4000 bits, energy consumption factor of amplification for the free space ($\in_{amp}$) is 0.0013 * 0.000000000001 and multi path ratio models($\in_{fs}$) is 10 * 0.000000000001.

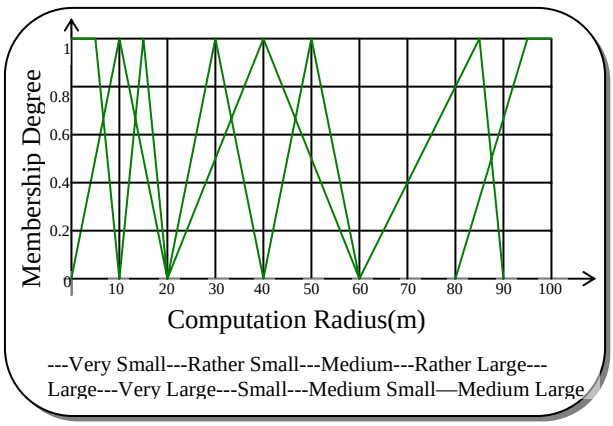


Fig 8. Fuzzy set for fuzzy output variable Computation Radius.

TABLE I: FUZZY IF-THEN MAPPING RULES FOR COMPUTATION RADIUS CALCULATION IN EAUCF

Distance to Base	Cluster Head Energy	Computation Radius
Close	Low	Very Small
Close	Medium	Small
Close	High	Rather Small
Medium	Low	Medium Small
Medium	Medium	Medium
Medium	High	Medium Large
Far	Low	Rather Large
Far	Medium	Large
Far	High	Very Large

IV. PUT FORWARD ALGORITHM

A novel perception has been commencing here as TSFL-LEACH using two-stage clustering approach and Fuzzy Logic. In this technique fuzzy membership value is calculated on the basis of distance between the cluster head and the base station. According to the fuzzy logic, the randomly placed cluster radius can be defined as ---

ALGORITHM 1: CLUSTERING ALGORITHM OF TSFL-LEACH PROTOCOL

1: Start
2: Initialize Energy of Every Node, Network Size, Number of Node, Number of Round(n), Cluster head percentage
3: Calculate Rcluster using fuzzy if-then mapping rules
4: CandidateCH(ID;Rcluster; resEnergy)
5: For round=1 to n
6: If round mod 2 = 0
7: if CR > d0/2 then
8: set ETx(k; d) = Eelec $\times$k + $\varepsilon_{amp} \times k \times d^4$
9: end if
10: End If
11: If round mod 2 != 0
12: if CR <= d0/2 then
13: set ETx(k; d) = Eelec $\times$ k + $\varepsilon_{amp} \times k \times d^2$
14: End if
15: End If
16: End for
17: Exit

Very Small---Rather Small---Medium---Rather Large---Large---Very Large---Small---Medium Small—Medium Large. In our proposed method we consider the local distance between the sensor nodes. In general clustering approach the distance d_0 for forming the cluster is not considered. In a cluster the distance between the sensor node and cluster head may be less or greater than the threshold distance (d_0). In that case a lot of energy is lost when the distance is slightly greater than the threshold distance. The losses can be saved by creating cluster depending on the maximum distance between the cluster head and other cluster sensor nodes. In our proposed method we

considered two different categories of clusters depending on the maximum distance between CH and sensors.

V. Performance Analysis

In this section, we present the results of the experiments that we have done to evaluate our algorithm. We compare our clustering algorithm TSFL-LEACH with three different algorithms, namely LEACH, LEACH-C and EEUC. We have implemented a WSN clustering simulator to evaluate our algorithm. This simulation tool is able to simulate LEACH, LEACH-C, EEUC and TSFL-LEACH for different WSN configurations. We have run several experiments with this tool to evaluate our algorithm. Experimental results show that our algorithm performs better than LEACH, TSFL-LEACH and EEUC.

TABLE II: Configuration Parameters

Parameter	Value
Network size	200 x 200
Base station location	100 x 270
Number of sensor nodes	200
Initial energy	0.5 J
Data packet size	4000 bits
$\in_{amp}$	0.0013*0.000000000001
$\in_{fs}$	10*0.000000000001

Figure 9 illustrates the circulation of the number of alive sensor nodes with respect to time for each algorithm. This figure perceptibly shows that alive rate is higher with respect to other algorithm. Before the time unit 500 the alive rate is 200 and it remains better with respect to other techniques until the time unit 1000.

Figure 10 illustrates the circulation of the number of dead sensor nodes with respect to time for each algorithm. This figure obviously shows that death rate is shorter with respect to other algorithm. Before the time unit 500 the death rate is almost zero and it remains better with respect to other techniques until the time unit 1000.

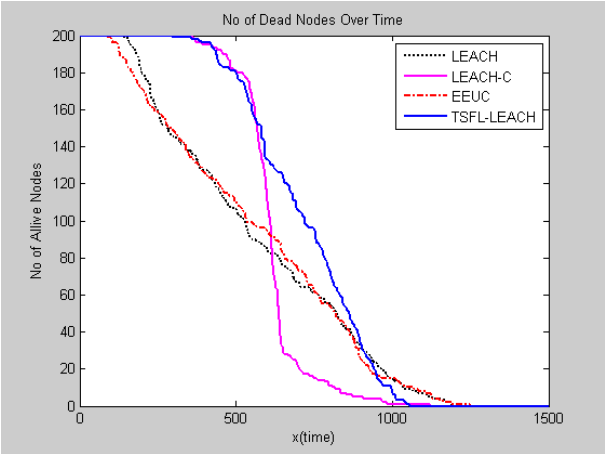


Fig 9. Distribution of alive sensor nodes according to the number of rounds for each algorithm.

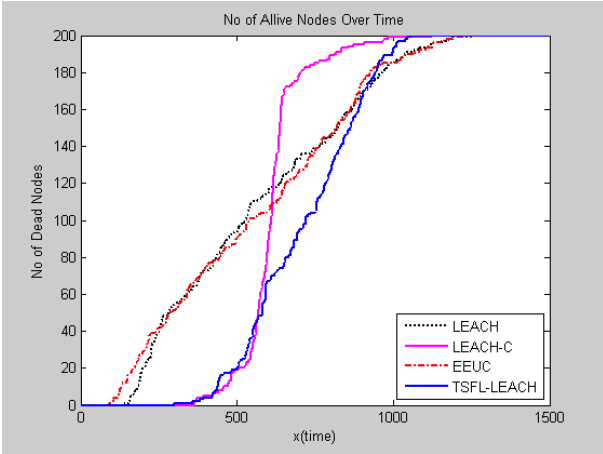


Fig 10. Distribution of alive sensor nodes according to the number of rounds for each algorithm.

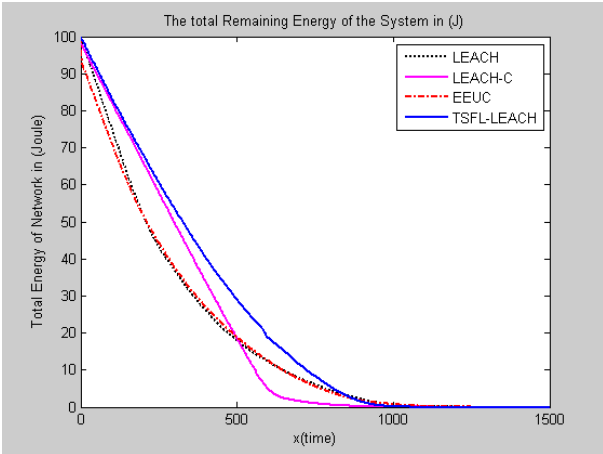


Fig 11. Distribution of alive sensor nodes according to the number of rounds for each algorithm.

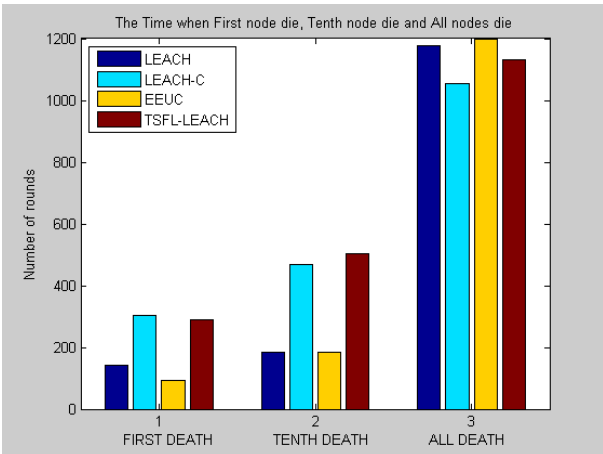


Fig 12. Values of FND and HNA metrics for each algorithm.

Figure 11 illustrates the allotment of the Total energy of the network in Joule with respect to time. This figure without a doubt shows that our proposed algorithm is more energy saving technique than the other algorithms, because sensor node energy remains more than the other technique.

Figure 12 illustrates the allotment of the Total energy of the network in Joule with respect to time. This figure without a doubt shows that our proposed algorithm is more energy saving technique than the other algorithms, because sensor node energy remains more than the other technique.

VI. Conclusion

In this paper, we have proposed a Two Stage Fuzzy Logic based Low Energy Adaptive Clustering Hierarchy, namely TSFL-LEACH. The key intention of our algorithm is to extend the life span of the WSN by consistently distributing the workload. To accomplish this target, we have typically paying attention on conveying appropriate cluster size depending on the distance among the sensor nodes. TSFL-LEACH regulates the cluster-head radius values bearing in mind the energy of the sensors and the distance to the base station from the sensor nodes. In general clustering technique the local sensor density is not considered. Consequently network traffic increases in the high compactness vicinity. Therefore, the sensor nodes on this locality die earlier. Our radius regulation method solves this problem.

We have shown that our proposed algorithm has an improved performance compared to LEACH, LEACH-C and EEUC considering the simulation results. In both of the scenarios, TSFL-LEACH sensor nodes start to die later than other algorithms. Moreover, total remaining energy level of TSFL-LEACH at a certain round is higher than all the other algorithms. Therefore, TSFL-LEACH is more energy efficient than other simulated clustering algorithms.

Considering these experiments, we can conclude that Two Stage Fuzzy Logic based Low Energy Adaptive Clustering Hierarchy is a stable and energy efficient clustering algorithm for WSNs. TSFL-LEACH algorithm is designed for the WSNs that have stationary sensor nodes. As a future work, our algorithm can be extended for handling mobile sensor nodes. In cluster-head computation, we only consider the residual energy of the tentative cluster heads. Some additional parameters such as node degree and local distance may also be considered as well.

References

[1] Hakan Bagci and Adnan Yazici, Senior Member, IEEE, "*An Energy Aware Fuzzy Unequal Clustering Algorithm for wireless Sensor Networks*," Fuzzy Systems (FUZZ), 2010 IEEE International Conference on Digital Object Identifier: 10.1109 /FUZZY.2010.5584580 Publication Year: 2010 , Page(s): 1 – 8.

[2] V. Loscrì, G. Morabito, S. Marano, "*A Two-Levels Hierarchy for Low-Energy Adaptive Clustering Hierarchy (TL-LEACH)*", Vehicular Technology Conference, 2005. VTC-2005-Fall. 2005 IEEE 62nd Volume: 3 Digital Object Identifier: 10.1109 /VETECF.2005.1558418 Publication Year: 2005, Page(s): 1809 - 1813

[3] Hakan Bagci and Adnan Yazici, Senior Member, IEEE, "*An Energy Aware Fuzzy Unequal Clustering Algorithm for wireless Sensor Networks*," Fuzzy Systems (FUZZ), 2010 IEEE International Conference on Digital Object Identifier: 10.1109 /FUZZY.2010.5584580 Publication Year: 2010 , Page(s): 1 – 8.

[4] Wendi Rabiner Heinzelman, Anantha Chandrakasan, and Hari Balakrishnan "*Energy-Efficient Communication Protocol forWireless Microsensor Networks*", Proceedings of the 33rd Hawaii International Conference on System Sciences – 2000.

[5] An Application-Specific Protocol Architecture for Wireless Microsensor Networks, "*An Application-Specific Protocol Architecture for Wireless Microsensor Networks*", Wireless Communications, IEEE Transactions on Volume:1, Issue:4 Digital Object Identifier: 10.1109/ TWC.2002.804190 Publication Year: 2002 , Page(s): 660 - 670

[6] Chengfa Li, Mao Ye, Guihai Chen, Jie Wu "*An Energy-Efficient Unequal Clustering Mechanism for Wireless Sensor Networks*", Wireless and Optical Communications Networks, 2007. WOCN '07. IFIP International Conference on Digital Object Identifier: 10.1109/WOCN.2007.4284192 Publication Year: 2007 , Page(s): 1 – 5.

[7] Zhou, Y. ; Hart, M. ; Vadgama, S. ; Rouz, A. Communications, 2007. ICC '07. " *A Hierarchical Clustering Method in Wireless Ad Hoc Sensor Networks* " Communications, 2007. ICC '07. IEEE International Conference onDigital Object Identifier:10.1109/ICC.2007.579 PublicationYear:2007, Page(s): 3503 – 3509.

[8] Xiaobo Chen ; Zhisheng Niu "A Randomly Delayed Clustering Method for Wireless Sensor Networks" Communications, 2006. ICC '06. IEEE International Conference on Volume:7 Digital Object Identifier: 10.1109/ICC.2006.255234 Publication Year: 2006 , Page(s): 3353 – 3358.

Performance Analysis of DS-CDMA Wireless Communication System with and without Diversity

Shuvashis Das Gupta, Sumaiya Shomaji, Farzana Islam, Tauhidul Hasan, Zisun Ahmed
Department of Electrical & Electronic Engineering, Ahsanullah University of Science & Technology
Tejgaon Industrial Area, Dhaka-1208, Bangladesh
Email: shuvoandshanto@gmail.com, sumaiyashomaji78@gmail.com, sadee_eee@live.com, zisunahmed@yahoo.com

***Abstract*— In this paper, performance of a novel analytic model of direct sequence code division multiple accesses (DS-CDMA) using diversity with maximal ratio combining (MRC) scheme is observed under the influence of multipath Rayleigh fading and multi-access interference (MAI) varying different system parameters and properties of PN sequences. The analysis presents a mathematical expression for bit error rate (BER) of the novel DS-CDMA model using diversity with MRC when signal to interference plus noise ratio (SINR) per bit follows a chi-square distribution. Performance results are evaluated in terms of average bit error rate (BER) at the output of MRC receiver by averaging the conditional BER over the probability density function (pdf) of signal to interference plus noise ratio (SINR). In addition to that the receiver sensitivity of the model is observed for different transmit and receive diversity in a proposed multiple input multiple output (MIMO) system model varying Rayleigh fading co-efficient. Variants of simulation result at various cases are also analyzed.**

***Index Terms*— DS-CDMA, BER, Diversity, MRC, SINR, MIMO**

I. Introduction

Mobile communications have huge impact on our everyday activities. For accommodating more users, more efficient use of bandwidth is a priority among cellular phone system operators [1]. Code division multiple accesses (CDMA) is one of the effective and efficient technologies is used in 2nd generation (2G) and 3rd generation (3G) wireless communication [2]. Code division multiple access (CDMA) is a promising technology for wireless mobile cellular communication for the reason of universal frequency reuse, soft-handoff and increased capacity [3]. It is a spread spectrum multiple access method. Wireless communication system always suffers multi user interference and different types of channel impairment like fading, Doppler shift and multipath propagation [4]. CDMA has a very unique feature to combat these effects by employing rake receiver by means of path diversity scheme [5].

CDMA in which the spread spectrum is achieved by directly multiplying the user's baseband signal with a high rate pseudorandom noise (PN) sequence series is referred to as direct sequence CDMA (DS-CDMA). In a DS-CDMA system, each user is assigned its own PN sequence, which is approximately orthogonal to all other sequences assigned to other users. Since the spreading PN sequences are not truly orthogonal, as a multiple access technology, CDMA is interference limited [6]. Multiple access interference (MAI), channel fading are some major limitation of DS-CDMA. Multiple access interference (MAI) is present in CDMA systems due to the nonzero cross-correlation between the different users' spreading codes [7].

Diversity combats channel fading by making use of more than one independently faded version of the transmitted signal. If several replicas of the signal carrying the same information are received over multiple channels that exhibit independent fading with comparable strength the chances of all the independently faded signal components experience deep fading simultaneously are greatly reduced. This will significantly improve transmission accuracy. There are several types of receiver diversity methods. In this paper we have used space diversity. For combination technique we have used maximal ratio combining (MRC). MRC achieves the best performance in linear combining [8]. Recent research has shown that combining DS-CDMA systems with multiple input multiple output (MIMO) techniques can achieve high gains in capacity, reliability and data transmission speed [9]. This is achieved by exploiting the spatial diversity made possible by multiple antennas at the transmitter and the receiver, allowing more degrees of freedom when the complex channel gains between different transmit and receive antenna pairs are sufficiently uncorrelated. In this paper we will also see MIMO DS-CDMA systems are also more robust to multiple access interference (MAI) than their single input single output (SISO) DS-CDMA counterparts.

The organization of the rest of this paper is made as follows: Section II introduces the system model of DS-CDMA under Rayleigh fading. Theoretical analysis are done in Section III. Results are presented in Section IV. Conclusions are achieved in Section V.

II. System Model

In DS-CDMA we directly multiply the user's baseband signal with a high rate PN sequence. The system model considered for this analysis is shown in Fig.1 and Fig.2 where Fig.1 is the functional block diagram of single DS-CDMA transmitter. Fig.2 represents the block diagram of DS-CDMA receiver using diversity with MRC technique.

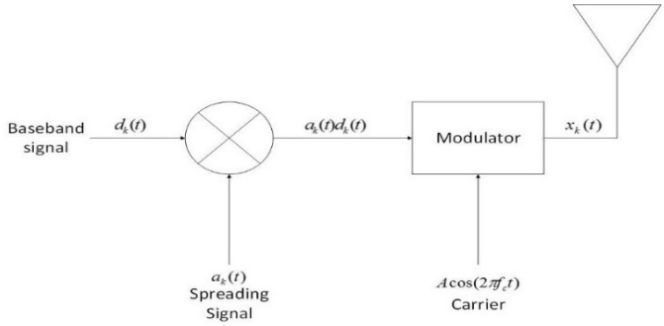


Fig.1. Block diagram of a DS-CDMA transmitter

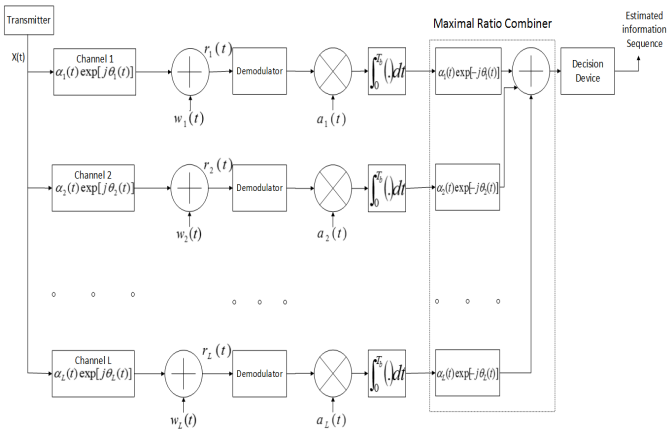


Fig. 2. Block diagram of a DS-CDMA receiver using diversity with MRC

The transmitted signal is

$$x_k(t) = A_c\left[\sum_i S_{k,i}\sum_{l=1}^{L} a_{k,l}\prod\left(\frac{t-iT_b-lT_c}{T_c}\right)\right]\cos(2\pi f_c t)$$

Where $a_{k,l} \in \{-1,+1\}$ is the l th chip of the binary PN sequence assigned to user k, $S_{k,i} \in \{-1,+1\}$ is the i th binary information bit, T_b is the information bit interval, T_c is the chip interval, corresponding to chip rate of $1/T_c$. Normally $T_b = LT_c$, where L is an integer. The principle of signal spreading is that L>>1 (i.e. $T_b >> T_c$). For simplicity we consider BPSK for the pass band modulation with coherent demodulation and PN sequences are periodic, with period L [6].

In receiving side, we consider the transmission of a digital modulated signal $x(t)$ over flat slow Rayleigh fading channels using coherent demodulation with L th order diversity. Using maximal ratio combining the decision variable for the kth transmitted symbol can be represented by

$$\vec{r}_k = \sum_{l=1}^{L}[\alpha_{lk}\exp(-j\theta_{lk})][\alpha_{lk}\exp(j\theta_{lk})\vec{x}_k + \vec{n}_{lk}].$$

Here, the demodulator in each channel is optimum for an AWGN channel therefore the output of the demodulator of the lth branch at the end of the kth symbol interval is $\alpha_{lk}\exp(j\theta_{lk})\vec{x}_k + \vec{n}_{lk}$, where $\alpha_{lk}\exp(j\theta_{lk})$ is the complex channel gain of the lth channel over the kth symbol interval, $\vec{x}_k$ is the vector representation of the transmitted signal over the kth symbol interval in the N-dimensional signal space and $\vec{n}_{lk} = (n_{lk,1}, n_{lk,2},..., n_{lk,N})$ is the corresponding vector representation of the noise component at the demodulation output. Each component in $\vec{n}_{lk}$ is a Gaussian random variable with zero mean and variance $N_0/2$ that is independent of any other noise component in the same diversity channel or in a different diversity channel.

III. Theoretical Analysis

This section contains three parts. First part consists of transmission performance analysis of a single cell DS-CDMA system using BPSK without diversity, second part consists of performance improvement by diversity over Rayleigh fading channels using maximal ratio combining. BER performance of DS-CDMA system with diversity has been proposed in third part and we have further expanded our analysis in a MIMO system in third part.

A. Transmission Performance Analysis in DS-CDMA system

Considering the forward link transmission of a single cell DS-CDMA system using BPSK, the output of the coherent demodulator is

$$\frac{1}{T_b}\int_0^{T_b} r(t)[a_1(t)\cos(2\pi f_c t + \theta_0)]dt$$

The output consists of three components:

1) The desired signal component, $(\sqrt{P_d/2})s_{1,1}$ which is a constant given $s_{1,1}$;
2) The noise component, which is a Gaussian random variable with zero mean and variance $N_0/4T_b$;
3) The intra cell interference component can be modeled as a Gaussian random variable with zero mean and variance $I_0 = \frac{(k-1)P_d}{2G_p}\varepsilon$; where $G_p = T_b/T_c$ is the processing gain and is the cross-correlation factor.

In summary, the demodulator output is a Gaussian random variable with mean $(\sqrt{P_d/2})s_{1,1}$ and variance $\frac{N_0}{4T_b} + \frac{(k-1)P_d}{2G_p}\varepsilon$ given that $s_{1,1}$ was sent. So, the probability of bit error is

$$P_b = Q\left(\sqrt{\frac{2E_0}{I_0 + N_0}}\right)$$

$$\Rightarrow P_b = Q\left(\frac{\sqrt{P_d/2}}{\sqrt{\frac{N_0}{4T_b}+\frac{(k-1)P_d}{2G_p}\varepsilon}}\right)$$

$$\therefore P_b = Q\left(\sqrt{\frac{2E_b}{N_o+2(k-1)\varepsilon(E_b/G_p)}}\right) \quad (1)$$

Where, $E_b = P_d T_b$ = the received signal bit energy.

B. Performance improvement with Diversity

For maximum ratio combining, as shown in fig. 2 the decision variable for the *k*th transmitted symbol can be represented by

$$\vec{r}_k = \sum_{l=1}^{L}[\alpha_{lk}\exp(-j\theta_{lk})][\alpha_{lk}\exp(j\theta_{lk})\vec{x}_k + \vec{n}_{lk}]$$

$$\Rightarrow \vec{r}_k = \left[\sum_{l=1}^{L}\alpha_{lk}^2\right]\vec{x}_k + \left[\sum_{l=1}^{L}\alpha_{lk}\exp(-j\theta_{lk})\vec{n}_{lk}\right]$$

$$\therefore \vec{r}_k = g_k\vec{x}_k + \vec{n}_{lk}$$

Where $g_k = \sum_{l=1}^{L}\alpha_{lk}^2$ and $\vec{n}_{lk} = \sum_{l=1}^{L}\alpha_{lk}\exp(-j\theta_{lk})\vec{n}_{lk}$; given the weighting gain $\alpha_{lk}\exp(-j\theta_{lk})$; $l = 1,2,.....L$. Each noise component $n_{k,n}$ is a Gaussian random variable with zero mean and variance $\sigma_{k,n}^2 = \frac{N_0}{2}\sum_{l=1}^{L}\alpha_{lk}^2$

For BPSK, N=1, the SNR per bit at the combiner for the *k*th transmitted symbol is

$$\gamma_k = \frac{[g_k x_k]^2}{2\sigma_{k,n}^2} = \frac{\left[\sum_{l=1}^{L}\alpha_{lk}^2\right]^2}{N_0\sum_{l=1}^{L}\alpha_{lk}^2} = \frac{E_b}{N_o}\sum_{l=1}^{L}\alpha_{lk}^2 \quad (2)$$

Where E_b/N_o is the SNR value for the AWGN channel with α_{lk} =1 and L=1. In a Rayleigh fading environment, the α_{lk} 's are iid Rayleigh random variables with parameter σ_α^2 . Therefore, γ_k follows a chi-square distribution with 2L degrees of freedom. Its pdf is given by

$$f_y(x) = \frac{x^{L-1}\exp(-x/\Gamma_c)}{(L-1)!\Gamma_c^L}, \quad x \geq 0 \quad (3)$$

Where, $\Gamma_c = 2\sigma_\alpha^2 E_b/N_0$ is the average SNR per bit each diversity channel. The analysis of the probability of bit error over a Rayleigh fading channel can be extended to the case with diversity, so that

$$P_b = \int_0^\infty P_{e|\gamma}(x) f_\gamma(x)dx \quad (4)$$

Where $P_{e|\gamma}(x)$ is the conditional probability of bit error given that the received SNR per bit is $\gamma = x$.

For coherent BPSK, we have $P_{e|\gamma}(x) = Q(\sqrt{2x})$ (5)

Substituting Equation (5) and (3) into equation (4), it can be derived that

$$P_b = \left(\frac{1-\mu}{2}\right)^L \sum_{l=0}^{L-1}\frac{(L-1+l)!}{(L-1)!\,l!}\left(\frac{1+\mu}{2}\right)^l \quad (6)$$

Where, $\mu = \sqrt{\frac{\Gamma_c}{1+\Gamma_c}}$ (7)

The BER transmission performance achieved by diversity with maximal ratio combining for coherent BPSK over Rayleigh fading channels.

C. Analysis of BER performance of DS-CDMA with diversity (using proposed model)

From equation (4), probability of bit error over a Rayleigh fading channel is

$$P_b = \int_0^\infty P_{e|\gamma}(x) f_\gamma(x)dx \quad ; \text{Where } P_{e|\gamma}(x) = Q(\sqrt{2x})$$

For DS-CDMA system using BPSK we propose that

$$x = \frac{2E_b}{N_o + 2(k-1)(E_b/G_p)\varepsilon}$$; Which is the SINR per bit.

So, probability of bit error of DS-CDMA system,

$$P_b' = \left(\frac{1-\mu}{2}\right)^L \sum_{l=0}^{L-1}\frac{(L-1+l)!}{(L-1)!\,l!}\left(\frac{1+\mu}{2}\right)^l \quad (8)$$

Where $\mu = \sqrt{\frac{\Gamma_c}{1+\Gamma_c}}$ and $\Gamma_c = 2\sigma_\alpha^2 E_b/N_0$.

To, analyze $N_r \times N_t$ MIMO system we consider h_{mn} is the channel response for *m* th transmitter and *n* th receiver.

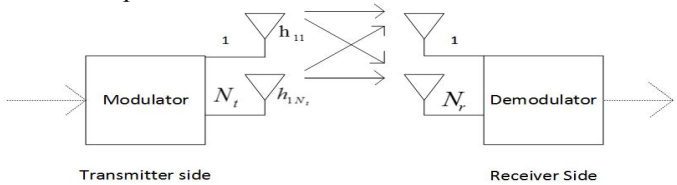


Fig.3. Proposed MIMO system for DS-CDMA wireless communication

Now,

$$(h_{11}^2+h_{12}^2+...+h_{1N_r}^2)+(h_{21}^2+h_{22}^2+...+h_{2N_r}^2)+...+(h_{N_t1}^2+h_{N_t2}^2+...+h_{N_tN_r}^2)$$

$$=\sum_{i=1}^{N_r}h_{1i}^2+\sum_{i=1}^{N_r}h_{2i}^2+...+\sum_{i=1}^{N_r}h_{N_ti}^2$$

Here, SINR per bit for each transmitter follows a chi-square distribution with $2N_r$ degrees of freedom. So, for N_t transmitters, SINR per bit also follow chi-square distribution and degrees of freedom for total combine model will be $2N_r \times N_t$ or $2L'$; where $L' = N_r \times N_t$.

Now, new expression for probability of bit error of DS-CDMA (MIMO) system over a Rayleigh fading channel is

$$P_b'' = \left(\frac{1-\mu}{2}\right)^{L'} \sum_{l=0}^{L'-1} \frac{(L'-1+l)!}{(L'-1)!\,l!} \left(\frac{1+\mu}{2}\right)^{l} \quad (9)$$

IV. Result And Discussion

Following theoretical analysis in Sec III, BER and other system performance results are numerically evaluated in this section.

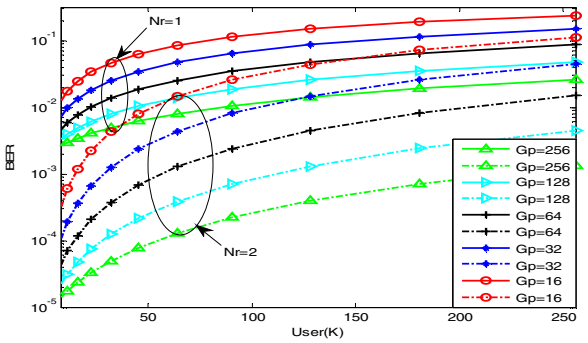


Fig. 4. Probability of bit error vs. Number of user (k) using diversity for different code length

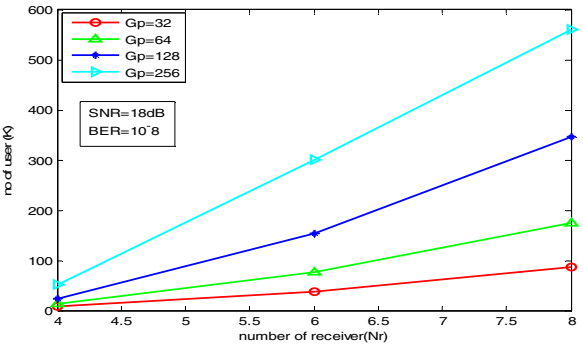


Fig. 5. Number of users (k) vs. number of receivers (N_r) for different code length

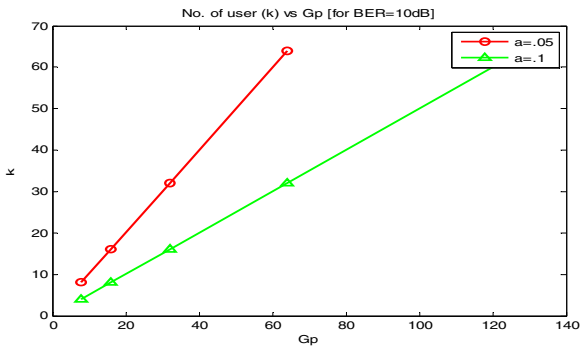


Fig. 6. Number of users (k) vs. chip length (G_p) for different cross-correlation factor

In fig. 4 Probability of bit error vs. Number of user is shown for different number of code length using diversity, it is observed that if number of users in a single cell DS-CDMA system using BPSK increase then the probability of bit error rate will increase because of multiple access interference (MAI), if code length increases then BER performance improves and if diversity technique with MRC applied in the system then BER performance improves significantly. Fig. 5 plots number of users vs. number of receivers for different code lengths and a fixed SNR and BER, here it is evident that higher number of users can be achieved by using higher code length and higher number of receivers. In fig. 6 number of users (k) vs. chip length for ε=0.05 and 0.1 is presented. It is noticed that as the value of cross-correlation factor, ε (i.e. a) decreases number of users using same chip length increase and increase in chip length increases the number of users in a DS-CDMA system.

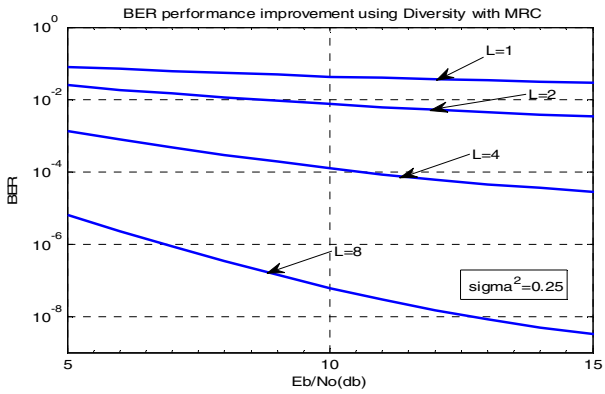


Fig. 7. BER performance improvement of DS-CDMA using Diversity with MRC, where σ_α^2=0.25

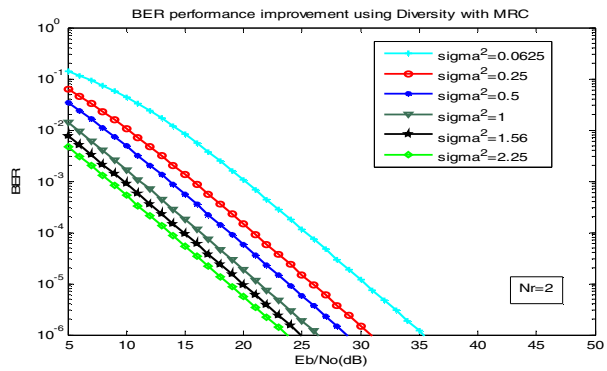


Fig. 8. BER performance improvement of DS-CDMA having different σ_α^2 using Diversity with MRC (N_r=2).

Using section III(C), fig. 7 presents the plots of bit error rate (BER) vs. SNR (E_b/N_0) in dB for a DS-CDMA wireless link using Diversity with MRC for σ_α^2 = 0.25, G_p = 128, number of users (k) = 64 and ε=0.1. It is observed that BER performance improves with the increase of the number of receivers (N_r=1, 2, 4 and 8) and there is a decrease in required E_b/N_0 at a given BER with increase in number of receivers.

In fig. 8 bit error rate (BER) vs. SNR, E_b/N_0 is shown for a DS-CDMA wireless link using Diversity with MRC for different Rayleigh fading co-efficient, σ_α^2. Here, G_p = 128, user (k) = 64, ε = 0.1 and number of receivers, N_r = 2. It is observed that there is an improvement in BER performance with increase in the Rayleigh fading co-efficient σ_α^2 for a given number of receivers.

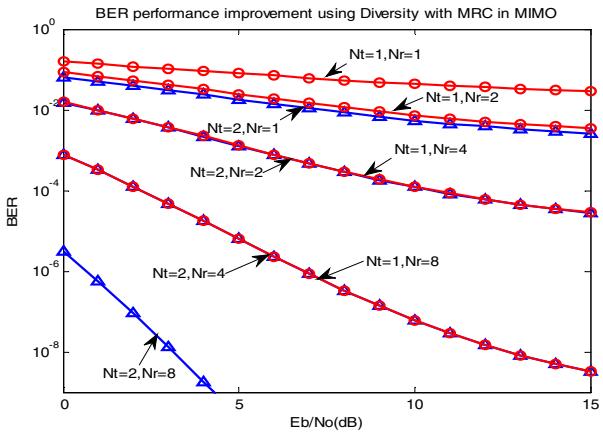


Fig. 9. BER performance improvement of DS-CDMA using Diversity with MRC in MIMO system.

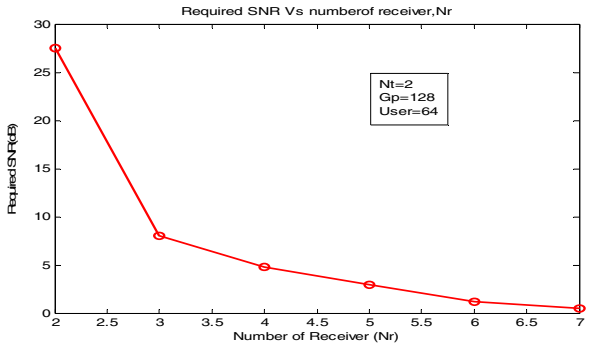


Fig. 10. Required SNR (dB) vs. number of receiver (N_r)

Fig. 9 presents the plots of BER vs. SNR (E_b/N_0) in dB for a DS-CDMA wireless link using Diversity with MRC for proposed MIMO system, where σ_α^2 = 0.5, G_p = 128, number of users (k) = 64 and ε = 0.1. By increasing the number of transmitters we can improve BER performance of the MIMO DS-CDMA system. It is noticed that there is decrease in required E_b/N_0 at a given BER (10^{-5}) with increase in number of receivers which is shown in fig.10 for a fixed number of transmitters, N_t =2.

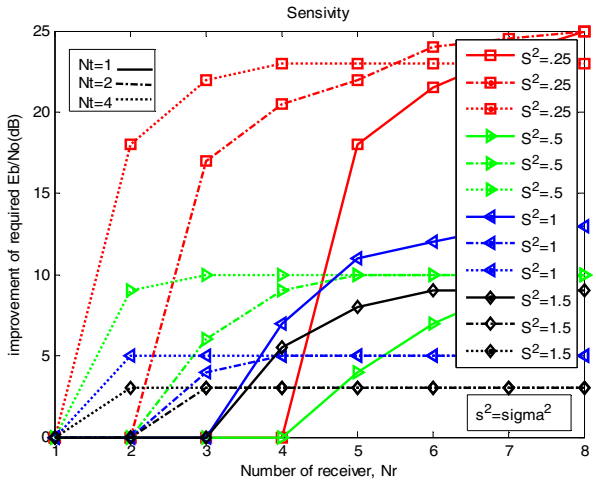


Fig. 11. Receiver sensitivity using Diversity with MRC in MIMO system for different σ_α^2

In fig. 11, the receiver sensitivity curves of the DS-CDMA using Diversity with MRC in proposed MIMO system is shown for different Rayleigh fading coefficient (σ_α^2). It is observed that the more σ_α^2 in the system, more the sensitivity curve shifts to left direction, which indicates less number of receivers can be used when σ_α^2 is higher. And if we increase number of transmitters (N_t) the receiver sensitivity curve also shifts to left direction.

V. Conclusion

This paper has demonstrated bit error rate (BER) performance of DS-CDMA wireless link with and without diversity. A mathematical expression for BER of DS-CDMA has been derived and performance is analyzed using diversity with MRC over Rayleigh fading channel. From proposed pdf of SINR bit error rate of receiver can be evaluated numerically. From numerical results it is found that if the number of users is increased in a single cell DS-CDMA system using BPSK then the probability of bit error rate (BER) will increase because of multiple access interference (MAI). Probability of bit error rate decreases when the code length is increased. Lower the value of cross-correlation factor, the probability of bit error rate will be lower. It has been noticed that, higher number of users using same code length can be achieved by decreasing the value of cross-correlation factor of PN sequences. There is an improvement in BER performance with increase in the Rayleigh fading co-efficient for a given number of receivers. It is noticed that there is decrease in required SNR at a given BER with increase in number of receivers. The receiver sensitivity of DS-CDMA wireless link improves with increase of number of receivers. By increasing Rayleigh fading coefficient in the DS-CDMA wireless system, the receiver sensitivity curve shifts to left direction, which indicates we can use less number of receiver when Rayleigh fading coefficient is higher. Some improvements in BER performance in proposed

MIMO wireless system are observed by increasing the number of transmitters. In future we expect to extend this analysis for Multicarrier DS-CDMA wireless communication system using rake receiver and the proper estimation of pdf of the chi-square distribution function for MIMO system.

ACKNOWLEDGMENT

We acknowledge the support of the department of Electrical & Electronic Engineering, Ahsanullah University of Science and Technology to carry out this work as a part of final year thesis of undergraduate programme.

REFERENCES

[1] Jean-Paul M.G. Linnartz – "Performance Analysis of Synchronous MC-CDMA in Mobile Rayleigh Channel with Both Delay and Doppler Spreads", IEEE transactions on vehicular technology, vol.50, no.6, November 2001.

[2] Md. Younus Ali, M. M. Ashiqur Rahman, S.M. Shamsul Alam, and Rajiv Kumar Mistry – "Evaluation of CDMA System Performance in Wireless Telecommunication", vol. 02, issue 02, IJCIT 2012, ISSN 2218-5224.

[3] Hui Liu, Member, IEEE, and Kemin Li, Student Member, IEEE "A Decorrelating RAKE Receiver for CDMA Communications Over Frequency-Selective Fading Channels ",IEEE Transactions On Communications, Vol. 47, No. 7, pp. 1036 – 1045, JULY 1999.

[4] N. Chan, "Multipath propagation effects on a CDMA cellular system,"IEEE Trans. Veh. Technol., vol. 43, pp. 848–855, Nov. 1994.

[5] Hassan, M.Z.; Tanzil, S.M.S.; Bhuiyan, T.A.; Hassan, M.M.; Majumder, S.P., "A combined analytical model on the bit error performance of a DS-CDMA wireless communication system in presence of Doppler shift, timing error and frequency offset," Electrical and Computer Engineering (ICECE), 2010 International Conference on , vol., no., pp.357,360, 18-20 Dec. 2010.

[6] Wireless Communications and Networking, chapter 6, Jon W.Mark, Weihua Zhuang, Prentice-Hall, 2003.

[7] S. Verd, Multiuser Detection, Cambridge University Press, Cambridge, UK, 1998.

[8] D.G. Brennan. Linear diversity combining techniques. Proceeding of the IRE, 47:1075-1102, June 1959.

[9] S.-M. Tseng, "Sequential detection for multiuser MIMO CDMA systems with single spreading code per user," IEEE Transactions on Wireless Communications, vol. 8, no. 7, pp. 3492–3497, 2009.

Designing A Press and Swipe Type Single Layered Bangla Soft Keyboard for Android Devices

Kaidul Islam and Bishnu Sarker
Department of Computer Science and Engineering
Khulna University of Engineering & Technology
Khulna, Bangladesh
ikaidul@yahoo.com, bishnukuet@gmail.com

***Abstract*—Development of text entry speed and accuracy is the key concern for designing and evaluating a soft keyboard. Bangla language has large collection of characters and it is hard to design a single layered fixed layout keyboard to provide high-performance in today's small mobile device. The traditional soft keyboards available in the market are multi-layered and menu based hierarchical keyboard which offer poor text entry speed. This paper describes the design, implementation and evaluation of a fixed layout single layered press and swipe type soft keyboard. Analysis of the comfort zone and movement of fingers, placement of characters and keys based on monograph, implementation on android platform and evaluation comparing with QWERTY keyboard have been described.**

Keywords—Android keyboard, Bangla keyboard, Keyboard layout, Single layered, press and swipe keyboard, monograph, soft keyboard.

I. INTRODUCTION

Numerous studies have been conducted on designing an efficient Bangla Soft Keyboard last several years. Bangla is one of the most spoken languages (ranked sixth) in the world with about 250 million native and 250 million total speakers [9]. And number of touch-screen device users is increasing exponentially. But after forty-two years of liberation war, a widely accepted and high-performed Bangla fixed layout keyboard has not yet been achieved. Bangla language has a large collection of characters and it is challenging to map all characters in a single screen of comparatively small devices. Most of the keyboards offer multi-layered layout results in low text entry speed and poor performance. So an efficient single layered soft keyboard needs to be introduced to fulfill better performance. Optimal keyboard layout designing is a potential research field and already many attempts have been taken to develop Bangla keyboard layout. There are about seventy letters including forty consonant, eleven vowels, ten consonant modifiers and ten digits. Some techniques can be followed to design an optimal layered keyboard to accommodate this large collection of letters. Monograph/Character frequency of Bangla language analyzed in a previous paper [5] is used to map most used character most comfortable movement zone of fingers. Five operations on a key – press and swipe in four directions are utilized to minimize the space of keyboard. At the end of the design phase, the keyboard has been implemented in Linux based Android operating system [16]. We have developed this keyboard based on two open source code by Steven Lin [14] and Gilbert [17]. We have facilitated all major features of modern android keyboard including prediction, learning dictionary, multi-tapping, quick-fixing and popup selection etc. After implementation, we have performed a user survey to verify its performance, text entry speed and accuracy. We have launched an experiment among 10 experienced users of our keyboard and QWERTY keyboard to ensure the proposed keyboard's feasibility by comparing text entry speed and error rates.

The paper is organized as follows: Section II describes the state of the art research and development over Bangla soft keyboard design in recent years. Section III delineates the proposed design methodology. The implementation detail is demonstrated in section IV. After that, experiments and evaluations are shown in section V. Finally section VI presents future plan and section VII draws the conclusion.

II. STATE OF THE ART RESEARCH

Designing a proper single layered layout to accommodate more than seventy characters into a small device screen is obviously a tough task keeping the usability and performance issues in mind.

A. Background Analysis & Related Work

It has been a while the first Bangla Keyboard named Munir Keyboard Layout [12] was designed in 1960. In the last several years, numerous researches on this field had been undertaken to design an efficient keyboard layout. There have been many endeavors made to design soft and hard Bangla keyboard based on data mining [4], finger and character frequency [5], Text-to-image replacement [3]. There are also some other remarkable researches on this field. Implementation of soft keyboards for Indic Languages is described in [1]. A multilingual touch-screen keyboard design is presented in [6]. Scott MacKenzie and Shawn X. Zhang presented design and evaluation of high-performance Soft Keyboard in [2]. There have been many hard keyboard layouts available for Bangla are Bijoy keyboard layout, Bangla inscript keyboard for Microsoft, Probhat keyboard layout [15]. Nowadays the most used and popular soft keyboard layout for is QWERTY (fig.1.) which we have considered as a reference to design our proposed layout.

B. Bangla QWERTY and Other Layouts

Bangla QWERTY layout is the modification of English QWERTY keeping similarities of utterance of characters [3]. This layout is the most used keyboard fixed layout in Bangla

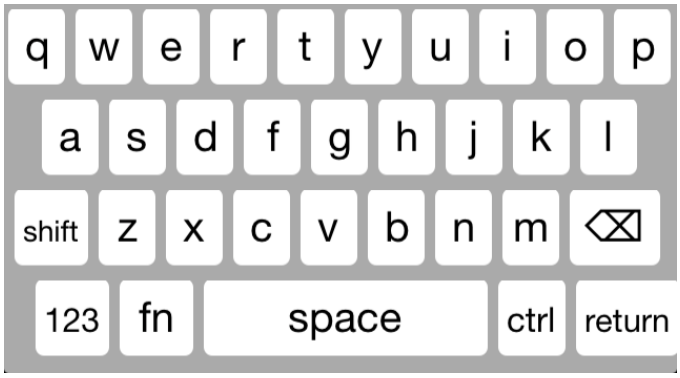


Fig. 1. Basic QWERTY Soft Keyboard Layout [15].

for touch-screen devices. This keyboard has 3 rows of keys consists of total 26 keys with 10, 9 and 7 keys in respectively. And this layout was first developed for English keyboard as English language has 26 letters only (excluding auxiliary symbols). This layout as a physical keyboard in computers and soft keyboard in smart devices is probably a "done deal" due to large existing world-wide user base and scientific significance of design. Fig. 1 Shows the Basic QWERTY Soft keyboard layout.

But this layout is not well-suited for Bangla language as Bangla has 70 frequently used characters which can't be accommodated properly into 26 keys. To accommodate such a large collection of characters, most keyboard layouts in Bangla including QWERTY are multi-layered and menu based, implies that more than one character are assigned in one key. Excluding digits and other auxiliary symbols about 60 characters are assigned in only 26 keys really affect the performance in a large extent. User has to press SHIFT key or pressing multiple times to type the inside characters. This really paves the way to fast text entry speed and movement-reaction time.

Some other mention-worthy keyboard layouts in Bangla are Bijoy, Panini Layout [18] which have the same issue of multi-layered and lack of performance. Designing a widely accepted layout for Bangla to fit all the keys properly in a tiny screen is really challenging.

III. DESIGN

Almost all the keyboard layouts in Bangla are in rectangular shape keeping edge corner's place quite blank. But previous research shows that a more practical square shape (or near square such as a 4 rows by several columns grid) is more practical for graphic design and still gives sufficient flexibility for optimization [19]. In practical product design, the keys on the edge unused by the alphabet letters can be filled with auxiliary characters, functions or comparatively less used characters. To manage some extra spaces to accommodate large number of keys, this paper demonstrates a square shaped layout.

To provide one separate key for each of the 71 keys will lead to a wide and long layout which is suitable for one hand typing and out of finger range. Also some devices are very little to accommodate them. So keyboard dimension should not exceed a suitable range.

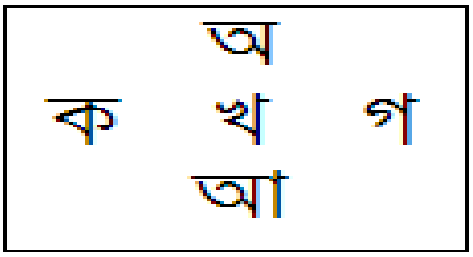


Fig. 2. Proposed Key Layout.

Fig. 3. Demonstration of Key functionality. Pressing key will type the middle character, swiping up will type upper letter, swiping left will type left letter, swiping right will type right character and swiping down will type down letter

Again creating tiny keys will create discomfort and lack of accuracy leads to low performance. Keys should have a minimum dimensions for comparatively big human finger.

Considering all these conditions, we have designed a comparatively bigger key to increase accuracy with five characters in one key. Fig. 2 and fig. 3 demonstrate the architecture of the proposed key layout.

This key will sense two basic operations – press (tap) and swipe (slide). When user press the key, the character at the middle will be typed. Swipe works in four ways – swipe up, swipe down, swipe left and swipe right. Swipe up will type the upper character and same as for the rest three swipe operations.

It is not necessary to press the key exactly upon the middle character or swiping exactly from/to the corresponding character, swiping or pressing anywhere on the button will work.

Accommodating 71 letters consisting of consonant, vowel and digits with few additional useful letters like space, ',', '.',

backspace, newline, symbols etc require at most 16 keys each consists of 5 letters with 4 rows and 4 columns.

To arrange the letters in a proper arrangement based on the character frequency – monograph, digraph and trigraph and developing a proper algorithm was another challenging task.

Fig. 4. Arranging letters keeping the possible adjacent ordering to provide user ease of decision making. Same color letters indicates adjacent five letters of a key.

So we undertook a user experiment among 30 advanced android users who frequently write blogs, update their Facebook status and do other stuffs on their device with Bangla soft keyboard. But the result brings no decision - Some users feel comfort to press the keys (not like to swipe the keys) and want the most frequency letters on the middle of keys. Some users want the opposite. Also there are different comfort zone for different users as some are left-handed and some are right-handed. Some feel annoyed to press the edge corner keys whereas some don't. And all this statistics have no majority to make a proper decision about the placement of the letters.

As our proposed layout is a single-layered layout and user can see all the letters at a glance, so arranging the letter without concerning about the letter frequency may arise very little efficiency lose. So we arrange the letters in our fashion with some empirical verification and monographic frequency shown in fig. 4. After placing all the letters into their corresponding keys, some possible optimization has been done.

A. Experiment on finding the more comfort zone of a key

As in our proposed layout there are five ways to type a letter on a key, there raised a question – which movement/operation is more comfortable for finger to type a letter. One can feel more comfortable in pressing than swiping. Or one can feel more comfortable in swiping upward than left side. So there can be 5! = 120 possible choices which would lead to a big experiment. But fortunately this experiment become small for us as all the users have agreed to prioritize pressing rather than swiping in any direction. So the experiment ended to 24 possible results. We have undertaken this survey among 20 correspondents who are advanced user of touch device. Most voted result was "middle > top > bottom > right > left".

There have been performed some good research on Bangla character frequency analyzing and Bangla monograph, digraph, trigraph analyzing [5].

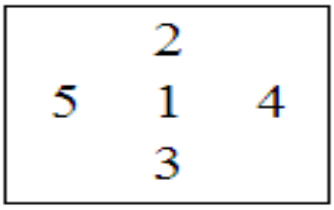


Fig. 5. User Comfort zone of Finger Movement. The lower the value, the best the zone is.

Bangla Character	Total Presence	Percentage	Bangla Character	Total Presence	Percentage	Bangla Character	Total Presence	Percentage
া	35578	12.96640	ী	3268	1.19102	ঃ	998	0.36372
ে	21436	7.81235	য	3226	1.17572	ফ	767	0.27953
র	18184	6.62716	জ	2812	1.02483	'	707	0.25767
ি	14326	5.22111	শ	2490	0.90748	ড়	554	0.20191
ন	13865	5.05310	ছ	2347	0.85536	ড	389	0.14177
ত	11951	4.35554	গ	2230	0.81272	ৎ	383	0.13958
ক	11910	4.34060	ও	2205	0.80361	ঠ	350	0.12756
ব	10465	3.81397	খ	1878	0.68444	ঘ	298	0.10861
ম	8751	3.18930	ং	1778	0.64799	ঞ	236	0.08601
ল	8490	3.09418	ধ	1655	0.60316	ঔ	230	0.08382
প	7979	2.90795	ট	1555	0.56672	ঢ	129	0.04701
হ	7278	2.65247	ষ	1502	0.54740	ঝ	110	0.04009
অ	6567	2.39334	থ	1438	0.52408	ঢ	89	0.03244
স	6532	2.38059	ভ	1214	0.44244	ঢ়	37	0.01348
দ	6483	2.36273	্	1184	0.43151	ঙ	17	0.00620
য	5200	1.89514	ণ	1169	0.42604	ং	9	0.00328
্	3454	1.25881	চ	1112	0.40527			

Fig. 6. Bangla character frequency and percentage [5].

Both researches are based on arbitrary documents on Bangla language. To determine the more frequent letters, we have used Bangla character frequency and percentage table from previous research undertaken by Md. Abdus Sattar et. al. [5]. Bangla character frequency and percentage table shown in fig. 6 enlist the Bangla characters along with their frequency and percentage.

We have rearranged the five letters of a key based on the frequency. A demo representation is depicted in fig. 5. A letter with greater frequency will be placed on a better finger movement comfort zone. Our algorithm for rearranging the letters of keys goes below.

B. Algorithm on Rearranging the the key label

1. Take all letters enlisted in fig. 6 into a stack which is already sorted in descending order of character frequency.
2. Pick a letter from the top of the stack and place it in the most comfortable movement zone available

(unmarked) so far on the corresponding key. If it is already in the desired place, then skip this step otherwise replace with the character of most comfortable movement zone.

3. Mark the letter obtained from top of the stack and the zone where it is moved.
4. Repeat step 2 & 3 until the stack is empty.

After placing all the keys in possible suitable positions and adding other necessary keys in suitable blank spaces our proposed layout looks like as shown in fig. 7:

Fig. 7. Mockup of Proposed Keyboard Layout.

IV. Implementation

Android is the world's most popular mobile platform. It is a Linux based operating system designed primarily for touch screen mobile devices such as Smartphone's and tablet computers. Android had a worldwide Smartphone market share of 75% during the third quarter of 2012, with 750 million devices activated in total and 1.5 million activations per day [16]. So it is obvious to design the keyboard on android platform so that people can easily and efficiently interact with android enabled device in Bangla. This will definitely inspire people to use Bangla more and more in their daily computing world and thus enrich the presence of Bangla language in computing and internet world. Also it is an open source platform with enriched developer support with full of resources on internet which helped us to develop it without any major problem.

A. Layout Overview in the android device

A view of the proposed layout both in portrait and landscape view are shown in fig. 8 and fig. 9 respectively. These figures demonstrate how it looks like when the proposed keyboard will be placed in real device environment for different screen orientation.

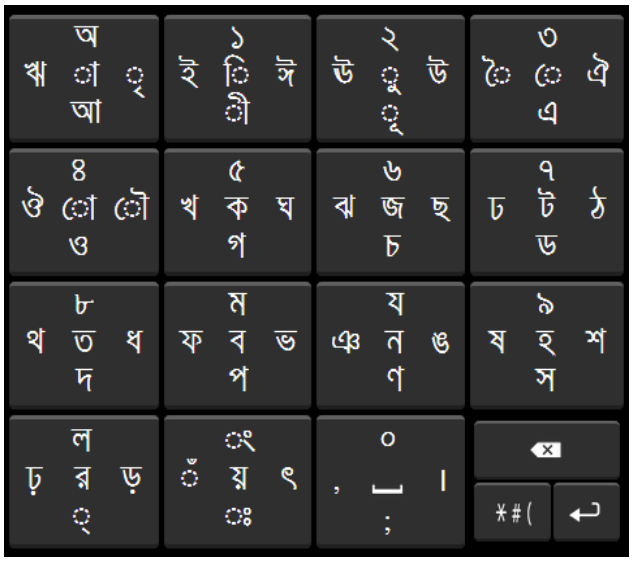


Fig. 8. Portrait view of the proposed keyboard layout.

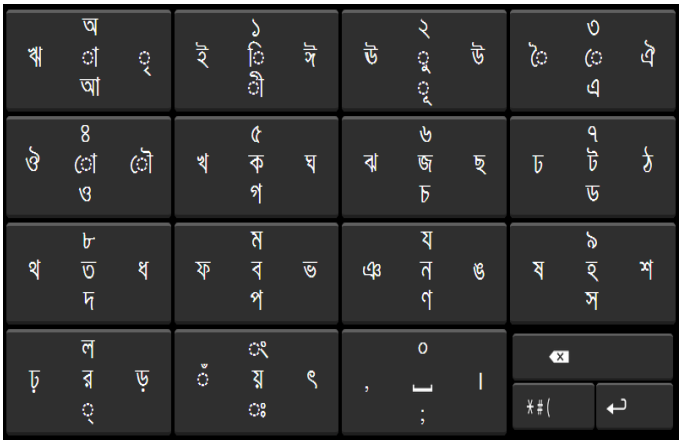


Fig. 9. Landscape view of the proposed layout.

B. Features

Developing the proposed soft keyboard requires knowledge about the android Input Method Editor (IME) related API, IME lifecycle. A deep analysis is encountered on open source code of Android Gingerbread Keyboard authored by Steven Lin [14] and S9 Keyboard by Gilbert [17]. So it relieves from developing it from scratch. The proposed press and swipe based Bangla soft keyboard is designed along with some promising features that tremendously facilitates the user experience and provides with flexibility.

a. Prediction on Candidate View

We have provided prediction view (Shown in Fig. 10) on key-press just like the other modern android keyboard for quickly picking desired text/words to increase text entry speed. Binary Dictionary is used for word list with native C/C++ code for faster manipulation.

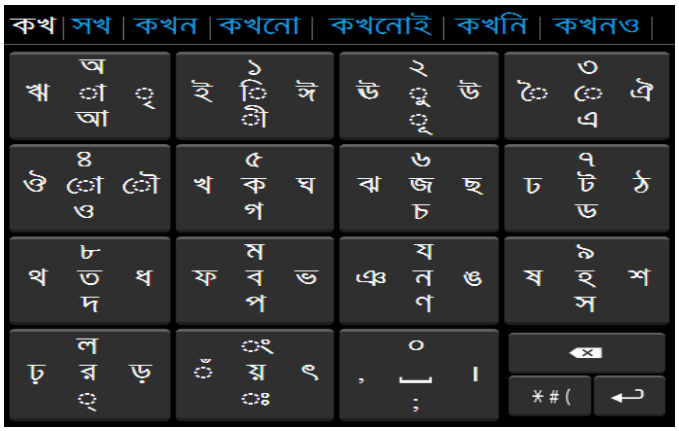


Fig. 10. Prediction View on Candidate View for better Performance.

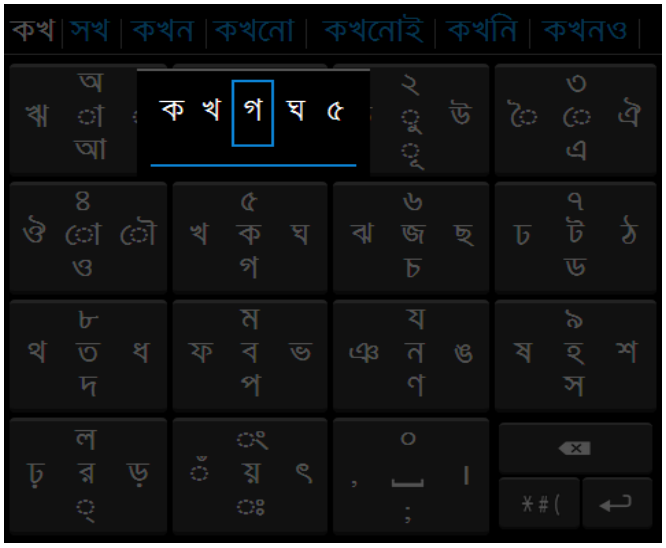


Fig. 11. Additional Popup Selection and Key Preview.

b. Quick Fixing

We have provided quick fixing functionality and flexibility to check it as ON or OFF for minimizing error rate.

c. Learning

User can add any word by long pressing on the text on the candidate view and add it to Binary Dictionary for later time availability of this word.

d. *Additional Popup Selection & Key Preview*

Users can long-press on any key and all the letters on this key will be popped up (shown in Fig. 11) to be picked up by users. This is not always time efficient but maybe handy sometimes. Also when user type or swipe any key, the typed key will be popped up as preview for making user sure that they have pressed or swiped the right letter they want to be typed

e. Multi-Tapping

User can press any key more than one time for getting typed the other letters except middle. When user presses any key two times, letter of up will be typed, three times pressing will type the bottom letter as so on.

V. Evaluation

It is always necessary to evaluate the usability and performance of a soft keyboard based on user feedback. There are many well-known Laws to evaluate and verify keyboard typing speed, movement latency from key to key, reaction time to make a decision for selecting a key. Fitts's Law [8] and Hick-Hyman Law [7] are two most popular methods to evaluate various language keyboard layouts. By calculating the time per character, WPM (word per minute) [1] can be calculated correctly. The performance of the proposed keyboard has been evaluated through a user study for comparing it with Bangla QWERTY keyboard on the basis of a given text.

A. User Experiment

To conduct the user study, user feedbacks are collected from 10 correspondents who are advanced android user for at least one year. All of them use Bangla keyboard to write blogs and interact with social networking sites. Most of them are used to type with various Bangla keyboards and familiar with their layouts. So users have been trained for at least 15 sessions each belongs to 10 minutes to make them familiar to proposed layout. Users are requested to type the following Bangla text with both keyboards: QWERTY and the proposed one.

ধন ধান্যে পুষ্পে ভরা আমাদের এই বসুন্ধরা,
তাহার মাঝে আছে দেশ এক সকল দেশের সেরা,
সে যে স্বপ্ন দিয়ে তৈরি সে যে স্মৃতি দিয়ে ঘেরা।

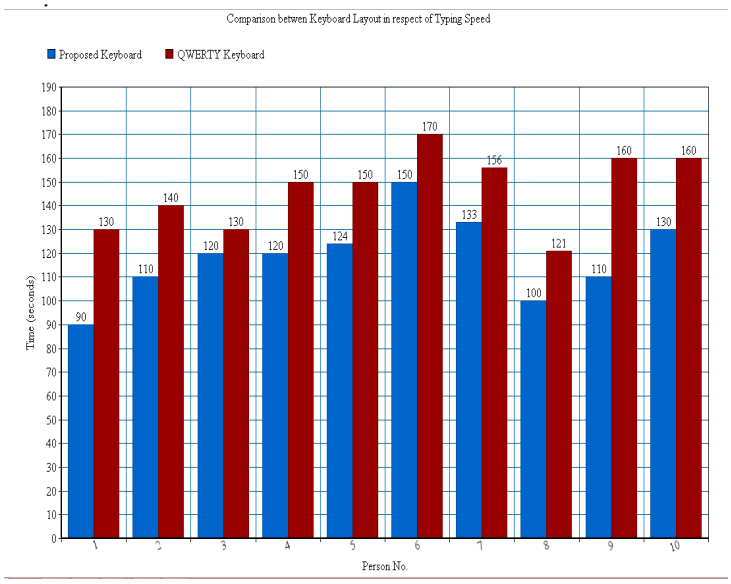


Fig. 12. Comparison between proposed Press and Swipe Layout and QWERTY Layout.

It is evident from the experimental outcome as shown in fig. 12 that proposed keyboard layout outperformed QWERTY keyboard layout as far typing speed is considered.

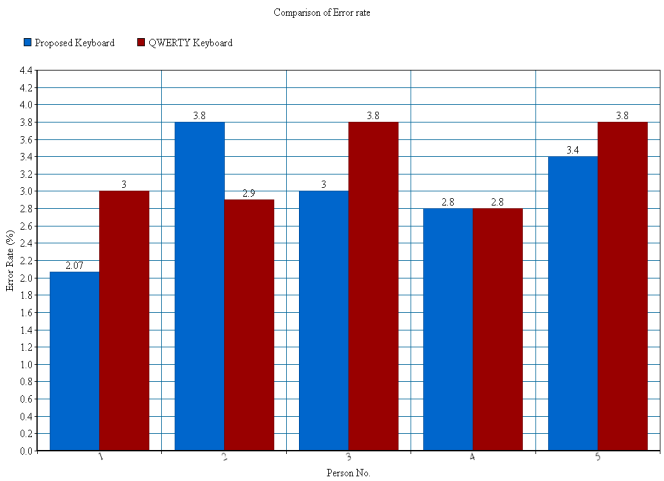


Fig. 13. Error Rate Comparison.

B. *Error Rates*

An error was recorded when the user entered character is different from the given/desired character. For the obvious reason the error rate while using the proposed keyboard initially is much in respect to QWERTY keyboard. This is because of the fact that the users are more acquainted with QWERT keyboard for long since. Users often pressed instead of swipe or swiped instead of press made those errors. But once the users will be more acquainted with the proposed layout the error rate may show a significant decrease. The experimental results are shown in figure 13.

VI. Future Plan

Designing a user friendly, interactive and best performed Bangla soft keyboard is a challenge. So it requires rigorous research and development works to meet the tradeoff between small display unit and text entry performance. Establishing a standard for evaluating the designed Bangla soft keyboard layout is also another big issue. There are many well-known laws to design and evaluate soft keyboard. The letter distribution and arrangement on the layout can be further developed by applying some well-known techniques like digraph-trigraph, data mining, machine learning techniques (genetic algorithm) and more user experiments. Our next research initiative would be to study and analyze the effect of applying the popular machine learning and data mining techniques to find out the most promising, adaptive, more interactive and a versatile layout for Bangla soft keyboard.

VII. Conclusion

The versatility of soft keyboards makes them an ideal research tool in seeking optimal layouts for Bangla Language. The methodology and layout presented in this paper can be easily reused in other language which has a large collection of letters and it is challenging to accommodate all characters in a single layered layout. This works has opened up the way to develop single layered layout with larger no. of letters. Android has launched swipe keyboard in their latest OS version Jellybean [10] which requires all the major letters in one layer. So android's Swipe keyboard functionality can be implemented on it by applying some modification. The proposed soft Bangla keyboard has been designed based on open source project and evaluated by undertaking user study. This novel design may bring about a shift in Bangla soft keyboard design that inspire single layer layout for languages having a large number of letters and characters and support both press and swipe for text entry.

References

[1] Albert Brouillette, "Implementation of Soft Keyboards for Indic Languages", Final Report, August 5, 2011.

[2] S. Mackenzie and S. X. Zhang, "The design and evaluation of a high performance soft keyboard," in proc. of the SIGCHI conference on Human factors in computing systems: the CHI is the limit, pp. 25–31, 1999.

[3] Hossain, S.I. et al, "Implementation of an Efficient Bangla Soft Keyboard with Text-To-Image Replacement Support", In proc. of 15th International Conference on Communication and Information technology, Chittagong, Bangladesh, 2012.

[4] Md. Hanif Seddiqui et. al, "An optimal Bangla Keyboard Layout", In proc. of 5th International Conference on Communication and Information Technology, Dhaka, Bangladesh, 2002.

[5] Md. Abdus Sattar, Md. Ameer Ali, and Mukaddim Pathan, "Development of An Optimal Bangla Keyboard Layout Based on Character And Fingering Frequency' National Conference on Computer Processing of Bangla, Dhaka, Bangladesh, 2004.

[6] Xiaojun Bi, Barton A. Smith, Shumin Zhai, "Multilingual Touchscreen Keyboard Design and Optimization", Human-Computer Interaction, Volume 27, Issue 4, 352-382. 2012.

[7] (2013) "Hick's law", Wikipedia.org [Online]. available: http://en.wikipedia.org/wiki/Hick's_law

[8] (2013) "Fitts's Law", Wikipedia.org.[Online]. available: http://en.wikipedia.org/wiki/Fitts's_law

[9] (2013) "Bengali Language", Wikipedia.org.[Online]. available: http://en.wikipedia.org/wiki/Bengali_language

[10] Johntanmi (2012), "Jelly Bean 4.2 Keyboard".[Online]. available: https://play.google.com/store/apps/details?id=jellybeankeyboard.f.g&hl=en

[11] Dan Macleod (2008), "10 Principles of Ergonomics" available: http://www.danmacleod.com/ErgoForYou/10_principles_of_ergonomics.htm

[12] "Munir Keyboard Leyout", [Online]. available: http://www.angelfire.com/tx/rezaul/munir.htm

[13] File:Bijoy Keyboard Image:jpg available at: http://en.wikipedia.org/wiki/File:Bijoy_Keyboard_image.jpg

[14] Steven Lin (2010), "Latin Keyboard from Android 2.3 aka. Gingerbread". [Online]. available: https://github.com/g1011999/Gingerbread-Keyboard

[15] "Ekusheyr Shadhinota 'Probhat' Layout". [Online]. available: http://ekushey.org/?page/probhat_layout

[16] (2013) "Android (operating system)", Wikimedia.org. [Online]. available : http://en.wikipedia.org/wiki/Android_(operating_system)

[17] Gilbert (2010), "An android input method that utilizes swipe motions to distinguish different characters on a single key".[Online]. available: https://github.com/gilbertl/S9

[18] (2013) "PaniniKeypad Bengali IME", Goole Play store Site [Online]. available: https://play.google.com/store/apps/details?id=com.paninikeypad.beng&hl=en

[19] Shumin Zhai, Michael Hunter, Barton A. Smith "Performance optimization of virtual keyboards", Human-Computer Interaction, pp.89-129, 2002.

E-Governance Using Social Network

A Model for Strong Democratic Environment in Bangladesh

Md. Amer Muntaq Ganim
Department of Computer Science and Engineering
Khulna University of Engineering & Technology
Khulna, Bangladesh
ganim09@yahoo.com

Md. Kamruzzaman
Department of Computer Science and Engineering
Khulna University of Engineering & Technology
Khulna, Bangladesh
mkzaman12@hotmail.com

***Abstract*—In recent years, investigating national incidents of Bangladesh it has been seen that social network played a vital role. Though social network sites like Facebook, Twitter, and Google+ are mainly used for connecting people all over the world but they are now being used for advertisements, news updates and for various important purposes of day to day happenings. Microblogging is getting more and more popular among educated and conscious people of Bangladesh. It is becoming another popular media which can provide entertainment and knowledge sharing about anything. Social Networking Sites (SNS) has become rendezvous of different people, different beliefs and interests to share and discuss. In this circumstance here we propose a model of e-governance using SNS for democratic Bangladesh perspective. Direct interaction between citizens and government can strengthen democracy. Accountability and transparency of politicians and government-workers can be ensured by flow of information to citizens of various developing activities. This paper proposes a model concerning these matters, in order to reduce corruption using a government controlled SNS.**

Keywords*—SNS (Social Network Sites), E-governance, Democracy, Transparency, E-Democracy, ICT (Information and Communication Technology), G2C (Government to Citizens).

I. INTRODUCTION

E-governance is the idea to connect root people with the government directly using web. There are several synonyms of e-governance including digital government, one-stop government and online government [3]. It is the use of modern information and technologies such as internet, local area networks and mobile phones by governments to improve effectiveness, efficiency for service delivery and to promote democracy. The objective of e-governance is to support and ease things for all parties. Its main purpose is to make a government more citizens centered. E-governance enables citizens to access government services and information through the use of internet or mobile phones. It brings transparency and reduces corruption. The government of Bangladesh is trying to build an e-government backbone. To keep pace with the modern world and to ensure a digitalized governance system, the government of Bangladesh identified four elements to give priority: e-governance, e-learning, e-commerce, e-citizen [1]. Already people of Bangladesh are enjoying some e-governance service. They are getting various results of public examination through mobile phones and websites. Many national institutions have their own websites. National web portal have been launched, so people can oversee the activities of government [15].

The term democracy originates from Greek word "dēmokratía" which means "rule of the people". Democracy is a form of government in which all eligible citizens have an equal say in the decisions that affects their lives. Democracy allows eligible citizens to participate equally, either directly or through elected representatives—in the proposal, development, and creation of laws. It encompasses social, economic and cultural conditions that enable the free and equal practice of political self-determination [16]. Modern political democracy is a system of governance in which rulers are held accountable for their actions in the public realm by citizens, acting indirectly through the competition and cooperation of their elected representatives [17]. But in the age of internet we learned about a new form of democracy, e-democracy. E-democracy is the use of information and communications technologies (ICT) and strategies by democratic sectors within the political process of local communities, states, regions, nations and on the global stages. Here the term "democratic sectors" means governments, elected persons, media, political parties, interested groups, civil society organizations, citizens and voters [18]. Bangladesh is practicing democracy more than a decade. Political parties also practicing e-democracy, they have their own websites and they are using social network to spread their news and holding election campaign. The use of internet-based social media programs that connects friends, family, classmates, customers and clients is known as social network. At present many social network sites are available such as Facebook, Twitter, Google+ etc. and they are playing a vital role. Social network now not only connects people and creates community but also has become a way of advertisement. Even political movements starts from social network. For example, we can say about Tahrir square in Egypt and Shahbagh movement in Bangladesh [19] [20]. In Bangladesh there are more than 3 million people who are connected to different social network, especially Facebook [13]. From the above discussion this paper proposes a model to develop a strong

democratic environment in Bangladesh by connecting e-governance, democracy and social network.

II. RELATED WORKS

It has been seen that a lot of researches have been done in the scope of e-governance. To strengthen the connection between government and citizens using new technology is the main objective of e-governance. In the late 1990´s the e-gov (Electronic Government or Electronic Governance) field arose as a new idea [3]. A brief discussion on the history of e-gov, current status of e-gov in 2004, content of the field as it appears in current research is done by author Thomas A. Horan [3]. In [12], authors present the concept of e-governance to describe and analyse local ICT policy and the role of various stakeholders in it and analyse the way European urban policymakers guide their city into the information age. Basic problems, major issues for implementing and acceptability of e-governance in India have been explained in the research by author Sanjay Kumar Dwivedi.

So there are a lot of researches on e-governance from history to implementing issues, problems, acceptability, new models for e-governance with different perspectives to different countries. E-governance using existing SNS technology for improving e-government services and proposing a model and guidelines for improving Serbian regional government entities social networks appearance in Serbia is the recent research on e-governance scope [5]. Bangladesh as a developing country and for coping with the technologies and challenging world there are also many researches on e-governance to improve Bangladesh.

Researching based on surveying of high ranked employees of various government organizations and politicians of oppositions and ruling parties of Bangladesh to find their ideas about E-governance authors proposed recommendation for improving e-governance in Bangladesh [6].

Launching of e-governance by Bangladesh government, opinions about e-governance in Bangladesh, capability of existing institutions to implement e-governance, challenges of practicing e-governance initiatives are discussed by author Mohammed Asaduzzaman [7].

In [8] authors have discussed on the issues that are the impediments of implementing e-governance and provide specific recommendations for implementing e-governance in the most feasible, cost-effective, and efficient manner in Bangladesh.

In [10], authors proposed a model of e-governance where main goal is to implement e-governance in a flexible and cost effective manner to suit with current infrastructure in Bangladesh to provide better information and service delivery of the government to the public; and hence promote democratic practices through public participation and consultation for urban and rural areas of Bangladesh.

In [9], authors researched on e-government for developing country and identified and analysed the challenges and opportunities of the initiative of e-government in developing countries.

In [11], authors present a proposal of a justifiable, feasible and effective framework for implementing e-governance by ensuring maximum utilization of existing infrastructure as well as providing a parallel initiation of the re-engineering process in Bangladesh.

From above discussion it is concluded that a many of researches have been done on e-governance in Bangladesh in different sectors. And our proposed model of e-governance will be new to Bangladesh that will help to strengthen democracy.

III. CONNECTIONS AMONG E-GOVERNANCE, DEMOCRACY AND SNS(SOCIAL NETWORKING SITES)

Government is the institution which decides different issues in national and international manner in order to develop a nation and to increase the standard of livelihood of the citizens. We can say a government is democratic when it is decided by the citizens of a country. Citizens elect their representative through direct vote in a democratic government system. Bangladesh, where at present democratic government system is going on over a decades but democratic environment is absent in most cases. People's participation, participating in decision making as a source of power is democracy. But in Bangladesh this aspect is not available. People's relation with the government is happened in two activities while casting vote for election and giving tax [18]. After being elected by the citizens, most of the representatives do not have any connection with the citizens. Many causes are responsible for this. Indifference of public representatives and unwillingness of citizens both are responsible for this. As an overpopulated country it seems impossible for an elected person to pay heed to everyone's problem. Bureaucratic complexity and lack of political moral keep away leaders from changing public needs and priorities [18]. For tedious activities in government offices people want to avoid them. So public problems and demands remains unknown. It hinders the main mantra of democracy. Democratic environment is greatly hampered in Bangladesh for the distance between government and citizens.

Government is trying various paths to connect with the people and to give them services. To make things easier government is using websites, mobile phones, social media to keep connect with the citizens, and usage of this digital process is known as e-governance. The main problem of this e-governance system is it is one-sided. Government keeps connection with the people but people can not connect directly with the government by e-governance. From Figure-1, we can see government passing information to citizens through digital media but citizen only works here as a receiver. They can not participate in any discussion to opine in this e-governance system.

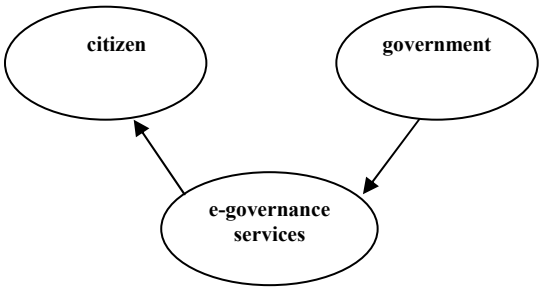


Fig. 1. Relation between citizens and government in existing e-governance services

To make citizen, an active participant through e-governance we propose a model of a Social Networking Site (SNS). SNS are widely used for connecting people with each other. From Figure-2, we can observe here a two sided communication and active participation of citizens where existing e-governance system are one-sided. So, by e-governance using a government controlled SNS it is possible to eradicate the distance of the government and the citizens where Bangladesh can create a more suitable democratic environment.

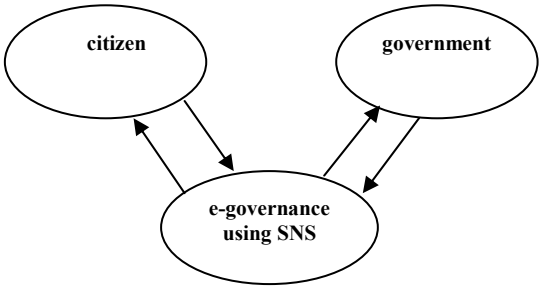


Fig. 2. Relationship between citizens and government using SNS as e-governance

IV. PROPOSED MODEL:SNS FOR GOVERNMENT

E-governance is composed of three major components are G2G (government to government), G2C (government to citizen) and G2B (government to Business) [2], [6]. G2C covers citizen's interaction with government [6]. Proposed model emphasizes on government to citizen issues and it is designed to eradicate the existing problems of democracy in Bangladesh. The model is essentially a SNS like Facebook, but we propose to develop a different SNS officially for e-governance in Bangladesh.

As Facebook is the largest and most popular existing SNS, government could use Facebook for e-governance but proposed model does not appreciate it because as this is an internal connection between the citizen and the government of Bangladesh. So the government has to build a new SNS which will act almost same as Facebook. By the term "will act almost same as Facebook", it expresses that the behaviour of the SNS will be like Facebook but the administrative and controlling will differ.

The model is based on two essential components: profiles and pages. Profiles represent citizens and pages represent government organizations, ministries and political parties etc. This model gives the framework and functionality of profiles and pages of this proposed SNS.

A. Profiles layout and functionalities

Proposed SNS comprises two types of profiles: citizen profiles and PR (Public Representative) profiles. Citizens of Bangladesh, having national identity card can have a citizen profile. As this SNS is government controlled and designed for discussing national issues, to avoid unwanted profile, verifying using national identity card seems feasible. A citizen profile can subscribe PR profiles or pages. When they would be logged in the SNS they can see posts given by their subscribed pages or PR profiles in the news feed. The term "news feed" defines the first page seen by a user after logging into a SNS. User having a citizen profile can comment on a post submitted by his/her subscribed pages or public representative profiles. User can subscribe new pages or PR profiles as well as can unsubscribe the previous subscriptions. Every user is eligible to post in the discussion forum attached with every page. The type of post can be texts, photos or videos. Subscriber of a particular page can also participate in a discussion post by other users. Every citizen profile must give some information about itself which can be seen by others.

Public representative profiles are provided by controlling organization of the SNS. Here public representatives are elected member of parliaments, mayor of city corporations, UP chairman and all types of elected members who are elected through direct election participated by the citizen of Bangladesh. The owner of a specific PR profile will change if necessary and it will be controlled by administration of the SNS. User having a PR profile will also has his/her own citizen profile. The main difference between a PR profile and a citizen profile is that a PR profile can post his opinion like a page can do where a citizen profile can only comment on posts. PR profile can also comment, subscribe to other PR profiles or pages. Another vital issue of a PR profile is that the user must share his election manifest and electoral affidavit in his/her profile. He/she will give his source of income, educational background and participation in various social works. User of a PR profile can share a message with the citizen through his profile to share his view and thinking.

B. Pages layout and functionalities

Every ministry of government and governmental institution will share their activities with the citizen through a

page. The managing administration of the SNS will provide the system of a page to the authority of a particular ministry or institution.

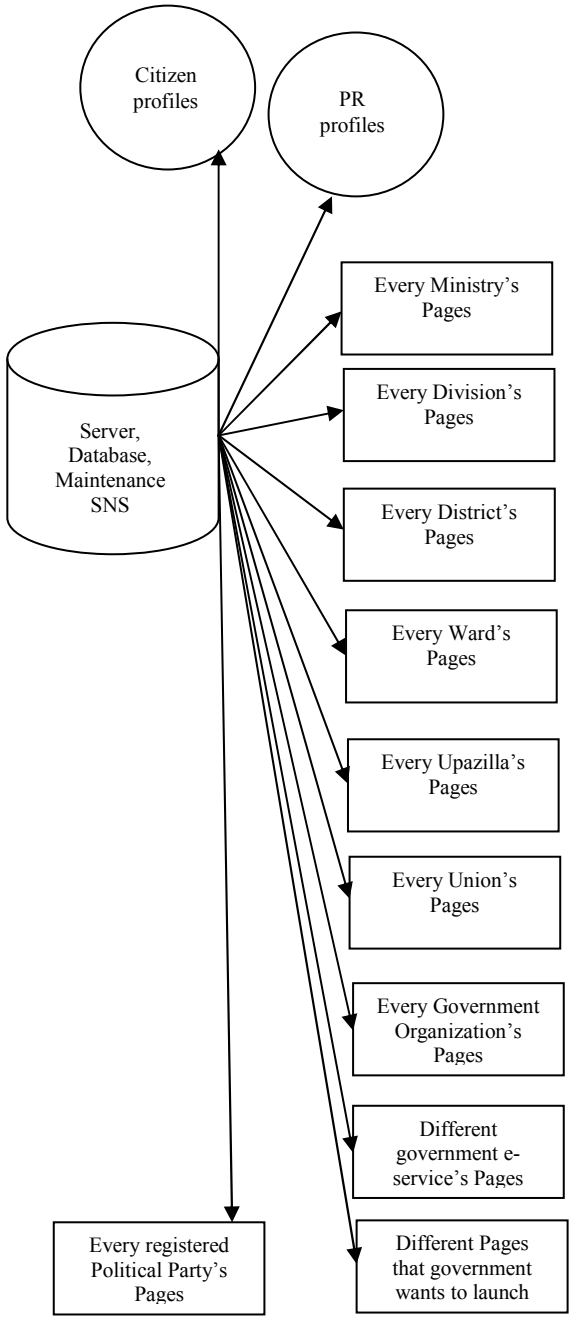


Fig. 3. Block diagram of the basic structure of proposed SNS.

Registered political parties will also get a page for their own purpose including election campaign. At the time of launching, the pages must contain some necessary information, such as in case of a ministry, it consists of name and profiles of responsible figures of the structure of the ministry including the minister and bureaucrats. It also provides important contacts so that anybody can access them while necessary. Governmental institutions like city corporations, divisional offices, district offices etc. will provide information contacts about the area such as contacts of police, fire brigade, tax office, passport etc. The pages will post their day to day happenings of their sector and flow the information to the citizens of Bangladesh. Government must ensure that every single information about public affairs come online. Citizens having citizen profile can comment on these posts and can express their joy, grief, opinions like other social networks. Pages hold the power to keep active this SNS. In every page, there is a discussion board, where citizens can discuss on a certain topic. Pages can launch public survey through the SNS.

Figure 3 shows the basic structure of the SNS that government will use. Circle shows the profiles and rectangle shows the pages of SNS.

V. A CASE STUDY

Let's assume, "Ministry of Finance" sanctions one billion taka for providing infrastructure of primary education and give it to "Ministry of Education". Both of the ministries will share this information with the citizen through their SNS page. Then when it will go to the divisional region, each division shares their information about this on their page. Following this rule, the very root level of institution will also provide their information. Citizens who subscribed these pages can comment on this action and can also opine in discussion forum. The anti-corruption commission and political parties keep close eyes on these events and can take necessary steps according to public opinion. By following this, citizens can know the information of an act to the very high to the very root level and transparency can be maintained easily. These processes are just like a tree, where the transparency between each child node and parent node can ensure the transparency of the whole system and information can reach to very root citizens and ensure a strong democratic environment where everybody can access information they want, can opine and thus participate e-democracy being a e-citizen.

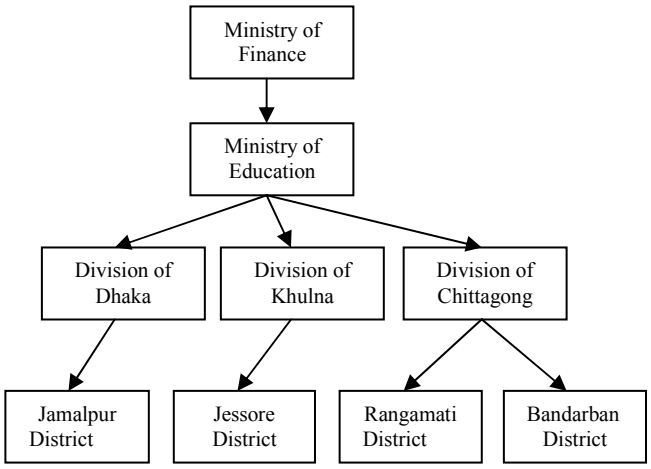
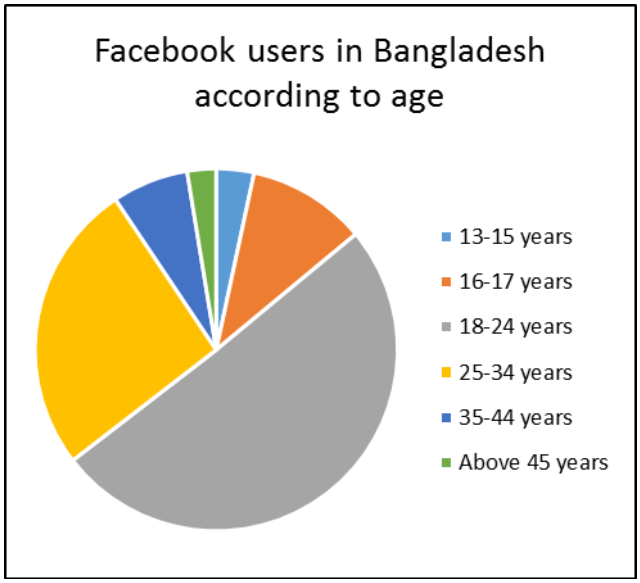


Fig. 4. Flow of information tree.

VI. ANALYSIS OF THE PROPOSED MODEL

Proposed model is fully related with the technology of SNS which may be handled and maintained by ICT (Information & Communication Technology) ministry. Bangladesh as developing country it will be very expensive for the infrastructures and maintenance cost. In the sense of corruption people behind this project who will maintain this will have to be honest persons. Issues of inadequate ICT infrastructures, unwillingness of government officials, power shortage, inefficient panning, Lack of awareness, absence of central powerful ICT wing, lower literacy rate, inadequate human resource, absence of effective legal framework, natural disasters, limited connectivity with high cost and low reliability, lack of reliable maintenance, lack of regulatory framework are the main impediments of implement the model [8]. These are some lacking of implementing SNS for Bangladesh, but it is possible if corruption can be handled by Anti-Corruption Commission of Bangladesh. Most of the educated and conscious people of Bangladesh are connected via most popular SNS Facebook. They posted their opinions and feelings in Bangla texts on different national issues using available Bangla keyboard via their computers, laptop, mobile phones etc. Investigating recent year's activities in Facebook, Bangladeshi users raised their voices via Facebook and still now they are discussing, criticizing, checking, talking, surveying, and sharing different national issues. Statistics (October 29, 2012) shows that Bangladesh is the 48th country in the world basis of the number of Facebook users [13]. There are 3.29 million Facebook users where internet users are 8,054,190 in Bangladesh in the end of 2012 [13]. Figure 5 shows the Facebook users in Bangladesh according to age [14]. So internet users and educated people are increasing in Bangladesh.



s

Fig. 5. Facebook users in Bangladesh according to age

From figure 5 it is seen that maximum users are 18-24 years. And perspectives of Bangladesh, in the age 18-24 maximum are students of Bangladesh [14]. Students are the future of the country but many of the students of Bangladesh spend many times in Facebook for entertainment but they are also conscious about their country. This time can be utilized besides entertainment if proposed model is implemented in Bangladesh. Student politics should not have to be banned rather it will be limited to online for developing purpose like checking out atrocity of government or other political parties etc. For election maximum time people don't know about the nominee who is bad or good for country. Even sometimes we have to vote without knowing him properly but e-governance can give them a profile where nominee can submit electoral affidavit or his information about him by which we can get a good knowledge about him though people never see him in front of them. And ICT and Anti-Corruption Commission will check the information that public representative has given to SNS is right or wrong. Using this people can get connected with each other and make unity among them for any development of Bangladesh and social media can coverage of activities of this SNS. Authors have made a survey among around 1100 Facebook users where 82% are students of different universities of Bangladesh 1% are students of MSc and Ph.D. in abroad and rest of them are job holders. The survey contains three questions about this model. Questions are in the TABLE 1. We created a form of these three questions in Google docs. Then, we shared the questions in Facebook publicly for mass participation.

TABLE 1. QUESTION USED IN SURVEY

Question NO.	Questions
1.	All of the developing information of Bangladesh that government is working on which is your right to know, if you get that in SNS will it be helpful for you? Yes or No?
2.	Discussion, criticizing, suggestions, making decisions on Government works is your rights as a citizen of democratic country. For this reason if government take step to develop any SNS then do you want to connect with that? Yes or No?
3.	People will more eager to discuss on national issues in Government SNS than existing SNS like Facebook or Twitter. Yes or No?

Figure 6 shows the survey result that 98.31% wants information about the development workings of government, and the vital thing is 89.83% people wants to connect with the proposed SNS for Bangladesh government but on the basis of popularity and entertainment 37.29% people thinks that people will not eager to discussing on national issues in proposed

SNS. But the majority thinks that people will be eager to discuss on national issues in proposed SNS.

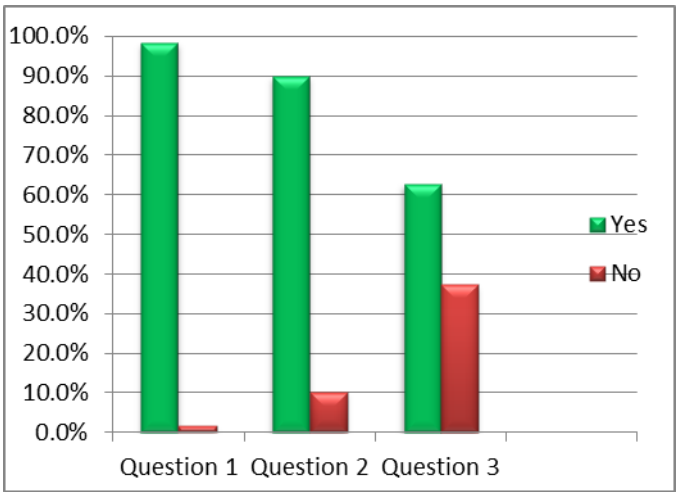


Fig. 6. Surveying among Facebook users basis of three questions about proposed model

VII. CONCLUSIONS

Although, the objectives and intensions of e-governance is good and it ensures involvement of the citizen with nation but due to some environmental, financial, economic and moral limitations it is causing inconvenience and citizens do not get the proper ramification. This paper proposed an overview of a model that will ensure the G2C and reduction of corruption in Bangladesh. Transparency of a democratic government is the main propaganda in proposed model but the fate of e-governance is fully depended on officials and public representative of Bangladesh. Citizens of Bangladesh becoming more educated and conscious and our survey clearly shows that they want to involve with the government activities directly by electronic services of government. In this era of technology Bangladesh as a developing country will have to cope with e-governance where our proposed model will help to build a better democratic country by ensuring transparency and reducing corruption.

ACKNOWLEDGMENT

The authors are thankful to those who have participated on the survey related to this paper and special regards goes to Dr. Faijul Amin for his thoughtful advices.

REFERENCES

[1] Journey Towards A Digital Bangladesh,Finance Division, Ministry of Fi nance, Government of the People's Republic of Bangladesh, June 2011.

[2] S. Prem Kumar, J. Keziya rani, e-Governance Applications for citizens-Issues and Framework, S. Prem Kumar et. al. / (IJCSE) International Journal on Computer Science and Engineering Vol. 02, No. 07, 2010, 2362-2365

[3] Å. Grönlund , T. A. Horan, INTRODUCING e-GOV: HISTORY, DEFINITIONS, AND ISSUES, Communications of the Association for Information Systems (Volume 15, 2004)713-72

[4] S. K. Dwivedi, A. K. Bharti, E-GOVERNANCE IN INDIA – PROBLEMS AND ACCEPTABILITY, Journal of Theoretical and Applied Information Technology,2005.

[5] A. M. Marković , A. Labus , Marko Vulić , B. Radenković , using social network for improving e- government services, 21st NISPAcee Annual Conferance, 2013.

[6] M. S. I. S., Assistant Professor, E-Governance in Bangladesh-Survey, Analysis and Proposed Recommendations, 19th Telecommunications forum TELFOR ,Serbia, Belgrade, November 22-24, 2011.

[7] M. Asaduzzaman , A. M. Rahman , N. J. Jinia, E-governance Initiatives in Bangladesh: Some Observations, Nepalese Journal of Public Policy and Governance, Vol. xxix, No.2, December, 2011.

[8] S. A. Ahsan Rajon, M. M. Rahman, Issues in implementing electronic governance: Bangladesh perspective, Computer and Information Technology (ICCIT), 2012 15th International Conference.

[9] Valentina (Dardha) Ndou, E-Government For Developing Countries: Oppurtunities And Challenges, EJISDC, 2004, 18,1, 4- 24.

[10] G. Z. Islam, M. I. Khan, M. A. B. Mazady, Developing a Model of E-governance for Urban and Rural Areas of Bangladesh, Proceedings of 11th International Conference on Computer and Information Technology (ICCIT 2008) 25-27 December, 2008, Khulna, Bangladesh.

[11] M. M. Rahman and S. A. Ahsan Rajon , An Effective Framework for Implementing Electronic Governance in Developing Countries: Bangladesh Perspective, Proceedings of 14th International Conference on Computer and Infonnation Technology (ICCIT 2011) 22-24 December, 2011, Dhaka, Bangladesh.

[12] A. v. d. Meer , W. v. Winden, E-governance in cities, Paper for the 42th congress of the European Regional Science Association 27 – 31 August, Dortmund, Germany.

[13] http://en.wikipedia.org/wiki/Facebook_statistics#Top_50_Countries_on_Facebook, 28 June, 2013.

[14] http://www.socialbakers.com/facebook-overview-statistics/, 28-6-2013.

[15] http://www.bangladesh.gov.bd/, 1 July 2013.

[16] http://en.wikipedia.org/wiki/Democracy, 1 July 2013.

[17] P. C. Schmitter and T. L. Karl, What is democracy is…and is not, Journal of Democracy,summer 1991, Article 25 page 114-120.

[18] S. Clift, E-democracy, E-governance and Public Net-Work, Open-Source-Jahrbuch 2004, hrsg. von Robert A. Gehring and Bernd Lutterbeck, Berlin 2004.

[19] http://shahbagmovement.com/, 1 July, 2013.

[20] http://en.wikipedia.org/wiki/Tahrir_Square, 1 July, 2013.

[21] https://docs.google.com/forms/d/1V9ySwPTu3qAyNS4b5wA-BYYWBXba3DVyAN8gXCbvowc/viewform, 1 July, 2013.

On the Impact of Virtual Environment in Trust Building: E-Commerce Perspective

S. A. Ahsan Rajon *and* Md. Mahbubur Rahman
Computer Science and Engineering Discipline, School of Science, Engineering and Technology
Khulna University, Khulna-9000, Bangladesh
ahsan.rajon@gmail.com, mahbubcse@yahoo.com

Abstract—**Virtual Reality is an important aspect of the state-of-art technologies. The emerging trend to integrate the highest level of interactivity in the computer-oriented activities like web pages has introduced a new dimension in various aspects of computer-oriented services. Electronic commerce (e-commerce) is such a sector, which may be greatly facilitated with the touch of interactive virtual reality or virtual environment. This research mainly aims to develop a trust establishment model for electronic commerce activities. This paper focuses on the usability of virtual reality and virtual environments in establishing trust in electronic commerce and related issues. We perform an extensive analysis on the current electronic commerce infrastructure and evaluate the effectiveness of virtual environment from the point of view of customer relationship management, trust-worthiness and customer satisfaction. An effective framework of integrating virtual reality with e-commerce is also presented in this paper. This paper also investigates the applicability of virtual reality and virtual environment for trust management in online shopping. The prime contribution of this paper is, it proposes an effective trust-building model to enhance the customer-seller relationship in e-commerce and provides a robust framework for the same. A survey has also been conducted evaluating the acceptability of proposed framework. Comprehensive recommendations on e-commerce-trust formation model in developing countries and its prospects and constraints are provided in this paper too.**

Keywords: Electronic Commerce, Virtual Reality, Virtual Environments, Customer Relationship Management, E-Commerce Trust, Customer Trust Management.)

I. INTRODUCTION

The emerging growth of computer-aided technology has ensured a smart and easy living throughout the world. From entertainment to military affairs, from chatting to research, everywhere there is a great variety of applications of computer and information technology. Though the rapid use of science and technology has provided a better way of quality living, the craving for reality in all the sectors involving technology has become a demand-of-time. This requirement of crossing the monotonous approach of computer oriented information technology has given birth to the new concern of virtual reality. With the sectors involving life and living, in other sectors of interest too, the applicability and usability of virtual reality has got rise. Immersive Virtual reality has become the prime tool for bringing technology much more closer, much more interactive and much more acceptable. E-Commerce is such a sector, which may be greatly enhanced with the touch of virtual reality. This enhancement may also facilitate the developing countries to cross a long way towards overall-development bridging the digital divide. In Customer Relationship Management (CRM) and in Customer Trust Building, the effect of virtual reality is unique.

The main contributions of this paper are, firstly, this paper presents an effective trust-formation scheme based on interactive virtual environment for electronic commerce. We also focus on the integration of virtual reality for representation of the products concerned with e-commerce framework. Secondly, it investigates various aspects of electronic commerce, trust and virtual reality which emphasizes on the factors of e-commerce trust formation. Moreover, whereas most of the existing schemes analyze the influence of any specific stream of catalyst, our proposed e-commerce trust establishment model considers about all possible factors influencing electronic commerce including customers' personal factors and social, cultural aspects as well as technical aspects. More importantly, our research aims to relate the three-dimensional component of trust, e-commerce and virtual construction schemes which is a new and a fresh step in e-commerce research. Finally, the acceptance of the proposed scheme has been evaluated through a survey.

II. TRUST BUILDING AND E-COMMERCE: RELATED RESEARCH

E-commerce is a process where the overall transactions may be carried out over internet. From the presentation of the products to getting response from buyer, everything is centered on computer-oriented activities. In order to convey the product to the consumer in a more appealing approach, the prime and essential requirement is trust building. The preliminary insight to the effectiveness of building trust in on-line shopping has been extensively covered in the literatures [1], [13], [14], [15].

Being an abstract concept, even though trust is an extremely important concern of e-commerce, it is really difficult to study. The definition of trust has rapidly varied from study to study, but one of the definitions that has prevailed was provided by Mayer et al. [13]. It defines trust as "the willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the trustor, irrespective of the agility to monitor or control that other party" [1]. The definition is general enough to encompass trust in all types of

operations and transactions and hence provides a benchmark for defining trust in e-commerce.

In trust building and trust management during online shopping (or e-commerce) activities there is a great necessity of integrating both the consumer psychology and the selling behavior conveyed through the internet trust development. In The Factor-Based Trust Building Model Mayer et al. [13] identified three characteristics of a trustee, ability, benevolence, and integrity.

The trustworthiness of a trustee is influenced by the key factors of trust which are ability (i.e. skills, degree of expertise, competencies, and perceived expertise of the trustee), Benevolence (degree to which a trustee is believed to want to do good to the consumer aside from an egocentric profit motive) and Integrity (refers to the consumer's perception that the trustee adheres to a set of principles that the consumer finds acceptable). These three key factors of trust is considered as the preliminary components in formation of initial trust in a web-store which is the root-activity of e-commerce [1].

The mostly adapted trust building models stated in [1], [5], [6], [7] and [8] describes hypothetical aspects in forming trust involving customer and e-merchant activities with approach of perceiveness and representation criteria. In [1], Shek et al. describes a trust building scheme where they describe that, enjoyment and perceived control coherently influence 'perceived usefulness' and 'perceived ease to use'. These two affects the trust of the consumer and consequently the trust-worthiness of the electronic commerce activity. The influenced degree of trust (positive or negative) either motivates the user to enhance the product reputation or discourages the user to have the product.

Tan and Thoen developed a generic model of trust where trust consists of two basic components: trust in the other party and trust in the control mechanisms used to ensure successful transactions. In The Process-Based Trust Building Model, McKnight et al. [14] describes another approach to specify trust formation. According to their model, initial trust is formed through three factors: firstly, individual disposition to trust secondly, institution-based trust, and lastly, cognitive-based trust.

In [23], an extensive study of trust has been provided by Grandison and Sloman. Pointing trust as an important aspect of decision making for Internet applications and particularly influencing factor for the specification of security policy, Grandison and Sloman provides a working definition of trust for Internet applications. They define trust as the firm belief in the competence of an entity to act dependably, securely and reliably within a specified context" (assuming dependability covers reliability and timeliness).

In [5], Ho and Wu employed case and survey methods to study the antecedents of customer satisfaction. They present five antecedents of customer satisfaction to be more appropriate for online shopping on the Internet which includes technological characteristics, logistical support, homepage presentation, information characteristics, and product characteristics. Among them homepage presentation is unique antecedent which has not existed in traditional marketing.

Building Online Trust through Socially Rich Web Interfaces has been focused in the literature by Hassanein and Head in [23]. They also explore the impact of various levels of socially-rich text and picture design elements on effect on online trust and its antecedents.

The relation between trust and various issues regarding web-site has been investigated in [8]. They conclude that, websites providing customers with enjoyment and perceived control leads to greater trust in the company itself through the customers' perceptions about the web site's usefulness and ease of use. In [20], Zarrai and Gharbi examines that which point the certification generates a cognitive chaining in the consumer, and to what extent it leads him to adopt approach behavior towards the commercial web sites.

In this paper, we integrate the technical factors relating customer trust, and specify virtual reality as a criterion for founding trust in online shopping and establish that, customer trust may be greatly modulated with the integration of virtual reality.

III. Integrating Virtual Environment and Virtual Reality in E-commerce

The prime objective of e-commerce is to establish a user-friendly, trust-worthy, effective and efficient approach of providing internet centered business transactions. In order to build trust among the consumer parties, there should have legible and reliable implications of the worthiness of credited information regarding product quality. The only means to ensure this in a credible and efficient way is to ensure virtual reality because; it is the only medium through which a person can explicitly interact with the concerned product and consequently gather reliable, accurate and exact information regarding the product in consideration. We may consider the cardinal virtue of virtual reality with e-commerce through the considerations of the following sector specific examples. The proposed model is greatly benefitted from the research of Koufaris and Hampton-Sosa [8].

In online purchase of electronic devices, virtual reality may serve a lot by providing a virtual three-dimensional glimpse of the product. As in most cases the prime issues relating the electronic devices are size, dimension, color and weight consumers are often interested to checkout those criteria-several of which may be supported using virtual reality. Using gestures users may be facilitated to have the views of those products from various elevations and in such a way the users may get a complete sense of the desired product.

For representations of mechanical devices to the consumers, the use of virtual reality may integrate an important role. The operational structure of the devices may be presented to the users only by using virtual reality [21]. Consumers may be given the opportunity to have the preliminary concept on the operating and controlling procedure of the device using avatars and embodied agents. This also introduces the applicability of e-learning [2], [3], [11] using virtual components. When the consumers will be given this opportunity to judge the operator-friendliness criteria of the device, it may put a more trustworthy impact on the consumer to choose the product.

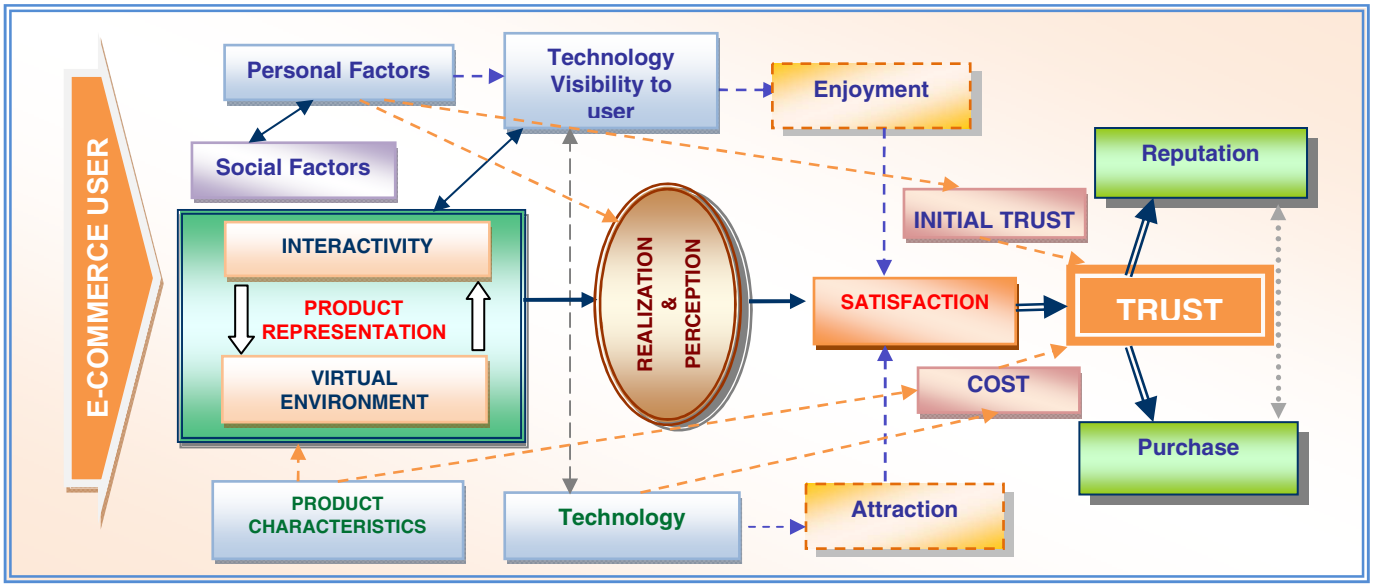


Fig. 1: Proposed Enhanced E-Commerce Trust-Building Model.

The representation of non-technical product or non-mechanical equipment may be made more attractive and appealing using virtual reality. The scopes of viewing the products from various elevations according to the consumers' interactivity make take him into closer interest to that product and thereby build a trust on that product as well as on the e-merchant. In obtaining customer satisfaction and consumer cognition, this may be really helpful.

Virtual reality may bring an insight to any product that clearly establishes a sense of trustworthiness. If we simply have a closer look at the housing trade, we easily find that, consumers may be given a complete view of the ongoing infrastructure of the concerned project over internet using virtual reality. This great power of virtual reality may also expand the scope of market and consequently strengthen the status of the e-organizations.

For online buying and selling of software products, virtual reality may serve a lot. In place of providing free trial or screenshots, if the users are given the opportunity to have a glimpse on the features of the software as well will complete scope to interact with that software using the virtual reality approach, it may be possible to attract the users easily by building the initial trust on the product as well as on the producing parties. This trust may undoubtedly enhance the customer-seller relationship and mobilize the e-commerce activities. Besides the software, other e-activities relating e-commerce; (*e.g.* e-books, e-labs, e-advertisements) may be turned most appealing with the interactive integration of virtual reality.

An inseparable part of e-commerce, which is being considered as the driving force of e-commerce too is e-entertainment. For the recent years, it is considered as the most profit making sector over internet. The first and foremost criteria of providing an interactive entertainment is virtual reality. It is the only medium which may ensure the maximum interaction with the entertaining objects like online games, online expeditions etc. The more interactivity will be added, the more users will be attracted and consequently the greater trust will be established among the users in comparison with the traditional approaches of simple multimedia.

Customer trust can be significantly increased when the company is perceived to have a good reputation. Again, reputation is defined as the degree in which people believe in both the company's honesty and concern towards its customers. That is, both trust and reputation are self-pinging. Consequently, the only way to build trust within customer is to provide the customers with maximum customization, customer-merchant interaction and customer reputation. Customization, which has been a buzzword in electronic commerce from its early adoption indicates the degree to which the electronic activity provides reputation on the customers interactivity and enables ease and frequent organization of the resources.

IV. Proposed E-Commerce Trust Establishment Model

In our proposed model, we integrate the factor of Virtual Reality which is modulated by interactivity to construct the trust formation model of electronic commerce. Moreover, we identify the potential of perception and realization on the product as the preliminary aspect of usefulness that is comprehended by the interactivity, user-friendliness and controlling criteria of the electronic commerce activity. Specifically saying, we integrate virtual reality as the means and of controlling and motivating the realization or perception criteria, which drives and derives the usefulness of the product in consideration. The effectiveness of the product is also motivated through the aspect of enjoyment that leads to greater trustworthiness of the product.

Whenever any user enters into any e-commerce site, the main aim of the web-site turns into attracting the user in such a manner that, the visitor perceives the concerned parties to be

enough trust-worthy to have transactions. This feeling of trust is naturally based on his/her personal factors which are formulated through the cultural and social ingredients. Moreover, the feeling of attraction by the visitor is also based on the technology visibility of the customer. Again, the technology visibility may be fluctuated through the product-representation style. Technology used to present the product is a prime factor to have quality representation. We propose to integrate virtual environment in presenting the product through internet with an aim to attract the consumer by providing virtual interactivity. Since this virtual environment is designed with consideration of cultural and social aspects of the local content, it will certainly attract the consumer to realize and perceive the features offered by the product. Whenever the perception and technology will have matchmaking, customer satisfaction may be achieved easily through his/her perceived enjoyment. This sense of satisfaction and his/her initial trust will certainly lead to trust for the concerned e-commerce activity. Hence, virtual environment will establish greater trust among the consumer.

Whenever the system will provide greater interactivity with virtual reality, the consumers will get the opportunity of higher control over the information system to get proper and comprehensive data. The user-friendliness criteria remain the same like existing models, which is influenced with the controlling infrastructure. We consider the user friendliness and control criteria as secondary inter-connected (long-dashed lines in fig. 1) component of trust building because, easier control results greater user friendliness whereas lack of control results lower user-friendliness. Again, the usefulness of any product may only be conveyed to the consumers when the 'usability information' is conveyed to them. This 'usability information' may be brought to the users only at that time when they would get the scope to interact with the system to get a closer inside at the system. This virtually interactive involvement with the products is raised from the controlling easiness i.e. user-friendliness. In the prolonged sense, the usefulness is realized though the resultant enjoyment in interacting with the system. The perceived usefulness forms trust within the consumer which initiates a customer to have the product or not.

The proposed e-commerce trust building model considers the human factors of the users like personal and social aspects, his or her technology visibility, initial trust on the product and the producer, previous experiences i.e. reputation on the product and the producer, enjoyment during visiting the site etc. Each and Every model of e-commerce trust including ours focus on the interrelations between the e-commerce users and the concerned products. In fact, trust itself is a human oriented factor which is obviously influenced by certain surrounding factors mostly by the social factors. This human factor like technology visibility, education, socio-cultural views greatly varies the sense of enjoyment and realization of the users. For example, a person with regular experiences of buying any product may have little regard for the enjoyment while for a novice user, enjoyment may be much more attractive factor to him/her and form some initial trust. Again the representation of the product is also varied with the difference in types of products. Whenever the product representation employs virtual reality and thus makes a very easier and flexible communication and interaction with the product over online, it is quite common that, there will have greater belief i.e. perception on the accuracy of information. For example, in the pre-VR phase of a mobile shop, when a person is provided the facility to get the various side views of the mobile phone, along with interacting with the menu, and thus makes the user think that he is interacting with the phone in the real world, it will attract the customer more profoundly rather than providing some static images of the cell phone or embedding a video. This multidimensional representation essentially initiates attraction and enjoyment which leads towards better satisfaction. Cost is another important factor from both seller and buyer end since, for a low cost product, any user is not supposed to spend more time for getting information on the product which leads towards necessity of adopting measures of rapidly attracting customer and supplying prompt information on the product. At the same time, for a product with greater price, the more a person is made to stay on the site, the greater is the satisfaction. To make a person stay more into website the best way is to provide exclusive control of the content to the user; i.e. incorporating maximum interactivity with the site for which the best approach is use of virtual reality for human computer interaction. As a consequence, initial trust is built on the product and the producer from the overall freedom and interactivity leading towards satisfaction.

V. Analysis of Proposed Trust Building Framework

A survey was conducted on 76 students of Department of Computer Science, Khulna Public College, Khulna, Bangladesh. Judgmental sampling method (a form of non probability sampling in which the population elements are purposely selected based on the judgment of the researcher) was employed to conduct the survey. Students enrolled in Computer Science and Information and Communication Technology courses were selected for this research since, they are supposed to possess basic knowledge on electronic commerce and internet as a part of their course-works and consequently the terminologies were comprehendible to the respondents. The students were also provided with a presentation on Virtual Reality before they take part into the survey.

Hypothesis 1 is developed assuming perceived benefit, initial company reputation and transactional accuracy as constant and non-repeating customers were considered. The main aim of this hypotheses group was to find out the key elements which initiate the attraction on the subject product and also on the seller.

Hypotheses group 2, was intended to identify the key means to representation of the products before the customers over online. In fact, various aspects of representational paradigms were taken into consideration as candidates.

Hypotheses group 3 tries to examine the perceptional factors of the users which affect in comprehending the representation of the product online and thus interacting with the total e-commerce system.

H1. *Primary Attraction i.e. Initial Trust*

H1.1: Personal and social factors primarily attract the customer.
H1.2: Cost of the product primarily attracts the customers.
H1.3: Enriched Product Presentation style (like interactivity and virtual reality) primarily attracts the e-commerce customer.

H2. *Representation*

H2.1: Integration of Virtual Reality ensures a complete representation of the product leads the user towards optimal perceived benefit.
H2.2: Simple and formal representation leads user towards optimal perceived benefit.
H2.3: Presenting initial reviews of other customers regarding products attracts the customer
H2.4: Scope of enjoyment through social network attracts the customer

H3. *Representational Perception*

H3.1: Users' Age influences the perception on represented product
H3.2: Technology visibility to the user influences the perception on represented product
H3.3: Education of the user influences the perception on represented product

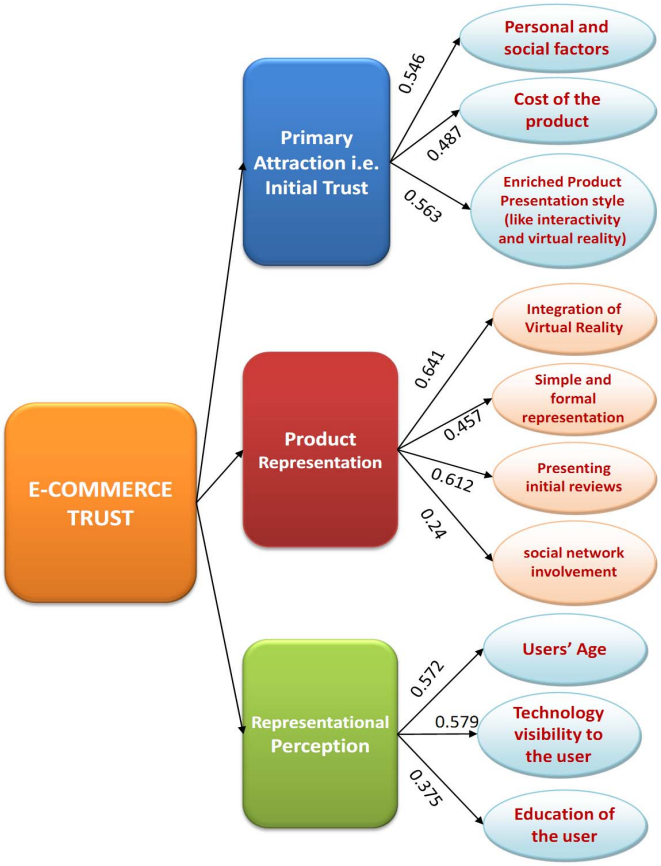


Fig. 2: Score of the survey results corresponding the Hypothesis.

For assessing the outcome of the survey, we considered the basic Likert Scale (l) (l ∈ {Strongly Agree, Agree, No comments, Disagree, Strongly Disagree}) with a bivariating factor of w; (w ∈ { 0.75, 0.5, 0, -0.5, -0.75}). The overall score of the survey was analyzed using the following equation:

$$s = \sum_{i=l_{\min}}^{l_{\max}} (f_i \times w_i)$$

That is, the frequency of respondents of each scale (i.e. agreed or disagreed) is multiplied with the corresponding weight factor (Suppose, 0.75 for strongly agreed, -0.75 for strongly disagreed etc.) and the summation results the overall score of the concerned hypothesis.

In figure 2, we have presented an overview of the findings.

Most of the respondents agreed that, enhanced product representation using virtual reality providing greater level of interactivity is much more important (having score 0.563) than cost of the product (0.487) while personal and social factors is appeared secondarily influential (0.546). That is, Hypothesis group 1 i.e. H1 is supported through H1.3.

For the second hypothesis group which is intended to assess the preferred medium and approach of product representation, the scores were significant for presentation of other users' review (0.612) and integration of virtual reality (0.641) while other scores were not so significant other than presenting simple, formal representation (0.457).

In perceiving the product representation, there is very close relationship of users' age and technology visibility with scores 0.572 and 0.579 respectively.

VI. Applications of Proposed Trust Building Framework: Hope for the Future

We have demonstrated that, integration of virtual reality may facilitate a lot in trust building. Again, trust building may influence the people a lot in increasing the reputation of any business process agent. Consequently, if it is possible to successfully integrate the two concepts, the developing countries may go a long way forward in the electronic commerce market establishing reputation on their products by gaining trust-worthiness of the consumers. Immersive virtual reality; being the most powerful, effective, efficient and successful medium of presenting the products before the consumers, may notice the worldwide consumers about the non-widely used products from the developing countries and hence establish their robust presence in the growing e-market. The main reason that places the developing third world countries at the backend of the emerging electronic commerce is their failure in presenting their product accurately and interactively in the e-market. In spite of having glorious culture and heritage which is artistically embedded in the products from these countries can not reach at the sight of the e-consumers. But, if the products especially the hand-made products with unique artistic styles can be conveyed to the concerned customers, the developing countries may get a touch of boosted economy.

It is to be noted that, the implementation of immersive virtual reality integrated electronic commerce should require specific modifications in the existing infrastructure. The prime obstacle in the way of implementing virtual reality is the speed and bandwidth of internet. But this great limitation may be partially handled by adapting efficient text compression schemes [25]. We also propose to employ immersive virtual

reality from the point of keeping the total model implementation able and cost-effective by avoiding the essentiality of head mounted display and other haptic devices and by only using cameras as the sensing and event-capturing devices.

The competitive world economy is now rushing towards developing trust among the consumers in order to strengthen the reputation on the product as well as on the e-merchant. The monotonous menu based navigation and multimedia-based non-interactive product presentation style is enlarging the distance between the consumer and seller. Because of the failure in trust-formation and non-interactiveness nature often, people are frustrated and diverged from online shopping. In order to change this situation, it is a must to provide some way to build trustworthiness among the consumers. The only effective way to establish trust is integration of virtual reality. Here a great question arises on the implementation and cost-effectiveness issue. The concepts of desktop virtual reality may serve a lot in these issues. A cost-effective framework of event capturing with low-cost display and reduced processing overhead may bring the trust model into practice. Consequently, the integration of virtual reality may bring the proposed trust model an effective one.

VII. Conclusion

The prime aspect of twenty-first century is technology. Standing at this glaring age of science, technology is now quite established. Though technology has ensured a faster, easier and better way of facilitated-living from the last decade, it has created frustration and monotonousness and has become unattractive to the consumers. The level of trust has also been greatly decreased in recent time. Only text or animation and even multimedia is now completely unable to form the sense of contentment or trust in online shopping. Consequently, it has become immensely necessary and to some extent essential to have some way to build or reconstruct trust among the consumers. In this paper, we have presented an effective and efficient trust-building model integrating immersive virtual reality, which possess the potential to build trustworthiness among the concerned parties and may enhance the quality of product presentation and uplift the overall electronic commerce environment. In this paper, we have also provided an extensive analysis on the integration of virtual reality with e-commerce to establish trust and hence enhance the quality of e-commerce.

References

[1] Sarah P. W. Shek, Choon-Ling Sia, Kai H. Lim, "A Preliminary Assessment of Different Trust Formation Models: The Effect of Third Party Endorsements on Online Shopping ", Proceedings of the *36th Hawaii IEEE International Conference on System Sciences* (HICSS'03), 2003.

[2] Jeremy N. Bailenson, Nick Yee, Jim Blascovich, Andrew C. Beall, Nicole Lundblad, Michael Jin , "The Use of Immersive Virtual Reality in the Learning Sciences: Digital Transformations of Teachers, Students, and Social Context", *Journal of The Learning Sciences*, Routledge, Taylor & Francis Group, LLC, 17:102–141, 2008.

[3] Bottentuit Junior, João Batista; Coutinho Clara, "Virtual Laboratories and M-Learning: learning with mobile devices", Proceedings of *International Milti-Conference on Society, Cybernetics and Informatics*, pg. 275-278, Orlando, EUA. Julho.2007.

[4] Laia Pujol, "Archaeology, museums and virtual reality", *Digit HVM Revista Digital dí Humanitats*, UOC, May 2004 ISSN: 1575 - 2275.

[5] Chin-Fu Ho and Wen-Hsiung Wu, "Antecedents of Customer Satisfaction on the Internet: An Empirical Study of Online Shopping", Proceedings of the *32nd Hawaii IEEE International Conference on System Sciences* – 1999.

[6] Nanda Kumar, Karl R. Lang and Qian Peng , "Consumer Search Behavior in Online Shopping Environments", Proceedings of the *38th Hawaii IEEE International Conference on System Sciences* – 2005.

[7] Kannan Srinivasan, Tridas Mukhopadhyay, Robert Kraut,Sara Kiesler, William Scherlis, "Customer Transaction Behavior on The Internet: An Empirical Study", Carnegie Mellon University.

[8] Marios Koufaris , William Hampton-Sosa , "Customer Trust Online: Examining The Role Of The Experience With The Web Site ", *Working Paper, CIS Working Paper Series*, 2002.

[9] Ezlika Ghazali, Dilip Mutum, "Exploratory study of buying fish online: Are Malaysians ready to adopt online grocery shopping?", *International Journal of Electronic Marketing and Retailing,* Vol. 1, No. 1, 2006, pp 67-82.

[10] Md. Shahidul Islam, M. Ismail Jabiullah, M. Lutfar Rahman, "M-Commerce Infrastructure Design and Development: Bangladesh Perspective", Proceedings of *7th International Conference on Computer And Information Technology*, 26-28 December 2004, Dhaka, Bangladesh

[11] Krista Johansen , Anita Dincesen , "Online Salmonella: Using Video and Interactivity to Teach Food Safety", *19th International Conference on Distance Teaching and Learning*, 2005

[12] Hunter G. Hoffman, Todd Richards, Barbaracoda, Anne Richards And Sam R. Sharar, "The Illusion of Presence in Immersive Virtual Reality during an fMRI Brain Scan" *Cyberpsychology and Behavior*, Volume 6, Number 2, Mary Ann Liebert, Inc., 2003,

[13] Mayer, R. C., Davis, J. H. and Schoorman, F. D. "In Integrative Model of Organizational Trust," *The Academy of Management Review,* 1995, 20(3), 709-734.

[14] McKnight, D. H., Cummings, L. L. and Chervany, N. L. "Initial Trust Formation in New Organizational Relationships," *Academy of Management Review*, 1998, 23(3), pp. 473-490.

[15] Lee, M.K.O. and Turban, E. "A Trust Model for Consumer Internet Shopping", *International Journal of Electronic Commerce*, 2001, 6(1) Page 75-91.

[16] Jamshid Beheshti, Andrew Large, Charles-Antoine Julian, "Designing A Virtual Reality Interface for Children's Web Portals", *Graduate School of Library & Information Studies*, McGill University

[17] Maria Roussou, "Immersive Interactive Virtual Reality and Informal Education", *Foundation of the Hellenic World*, pp. 1-5.

[18] Fred Brooks, "Is There Any Real Virtue In Virtual Reality?", University of North Carolina, Chapel Hill, April 21, 1998,

[19] Frederick P. Brooks, "What's Real About Virtual Reality?", University of North Carolina at Chapel Hill, IEEE Computer Graphics and Applications, 1999

[20] Rihab Zarrai and Jamel-Eddine Gharbi, "Perceptual and Personal Determinants Of Trust On A Commercial Web Site", 2008

[21] Jeff Rickel and W. Lewis Johnson, "Animated Agents for Procedural Training in Virtual Reality: Perception, Cognition ,and Motor Control", *Applied Artificial Intelligence* 13:343-382, 1999.

[22] Hee-Woong Kim , Yunjie Xu, Joon Koh, "A Comparison of Online Trust Building Factors between Potential Customers and Repeat Customers", *Journal of the Association for Information Systems*, Vol. 5 No. 10, pp.392-420, October 2004

[23] Khaled S. Hassanein and Milena M. Head , "Building Online Trust through Socially Rich Web Interfaces", pp. 15-22, 2004

[24] Arun Sukumar, David Edgar , "E-Business, SMEs and Risks: Towards a Research Agenda", *International Journal of Management Innovation Systems,* ISSN 1943-1384, Vol. 1, No. 2, 2009.

[25] S. A. Ahsan Rajon and Md. Rafiqul Islam, "An Enhanced Short Text Compression Scheme for Smart Devices", *Journal of Computers, Academy Publisher*, Vol. 5, No. 1, January 2010. doi:10.4304/jcp.5.1.49-58.

Radius Particle Swarm Optimization for Resource Constrained Project Scheduling Problem

Mana Anantathanvit
Faculty of Information Science and Technology
Mahanakorn University of Technology
Bangkok, Thailand
m.anantathanavit@outlook.co.th

Mud-Armeen Munlin
Faculty of Information Science and Technology
Mahanakorn University of Technology
Bangkok, Thailand
mmunlin@gmail.com

***Abstract*— The Resource-Constrained Project Scheduling Problem (RCPSP) is a classical well-known and NP-hard problem which includes the resource and precedence constraints that has been applied to many applications. This paper proposes the Radius Particle Swarm Optimization (RPSO) to solve the RCPSP. It extends the Particle Swarm Optimization (PSO) by regrouping the agent particles within the appropriate radius of the circle. It initializes the group of particles, calculate the fitness function, and find the best particle in that group. It is then applied to the RCPSP to find the best scheduling. The proposed method is tested against the PSPLIB standard datasets. The results show that the R-PSO gives better average makespan and standard deviation than traditional PSO.**

Keywords*— *Resource constrained Project Scheduling Problem (RCPSP), Particle Swarm Optimization (PSO), Radius Particle Swarm Optimization (R-PSO)

I. INTRODUCTION

The resource-constrained project scheduling problem (RCPSP) is a classical problem in the group of scheduling and NP-Hard problem [1]. The RCPSP is a part of the current business, such as production planning, construction engineering, machine operation and software development. The current business deals with process, resource allocation and material requirements in order to complete the task on time. As a result, the business can estimate the project duration and the cost consideration relate to the resource constraints. The goals are to produce the scheduling as to continuously improve the value-added benefits of the operations.

The RCPSP is a complex problem which involves resources and precedence constraints. Many studies have solved the RCPSP by applying the methods and the mechanisms to find the better makespan and minimize the time of the scheduling. These methods include exact, heuristic and meta-heuristic procedures [1,2,3]. Several methods have been proposed to solve the RCPSP. Hapka and Slowinski [4] present generalization of the known priority heuristic method and create priority list by fuzzy set. Lim et al. [5] presented new concept of float and critical activity which developed the schedule without activity start and activity end and suggest several solutions using the heuristics.

Many papers solve the RCPSP by applying the meta-heuristics and the heuristics based schemes, such as Bee Algorithm [6]. Mobini et al. [7] present Artificial Immune Algorithm (AIA) using a meta-heuristic algorithm by developing the schedule with the serial generation schemes (SGS). Agarwal et al. [8] present a hybrid genetic algorithm (GA) that use neural network (NN) to pick the best solution from the GA. Montoya-Torres et al. [9] present the GA algorithm for the RCPSP that used a multi-array Object-Oriented Model which set the priority and a delay time in the chromosome. Hong et al. [10] present the Particle Swarm Optimization algorithm (PSO) to solve the RCPSP by developing the schedule with the parallel generation schemes (PGS). These methods are tested against the standard dataset of the project scheduling problem library (PSPLIB).

In this paper, we present the Radius Particle Swarm Optimization (R-PSO) to solve the RCPSP using makespan and time as the objective function. We extend the traditional PSO algorithm by regrouping the particles within the appropriate radius and compute the agent or local best particle of each group. These candidate particles are employed to iteratively find the global best particle or the best solution under the given objective functions.

II. PARTICLE SWARM OPTIMIZATION ALGORITHM(PSO)

The PSO is a population-based stochastic algorithm [5]. It initializes a population with random positions and search for the best position under the given fitness function. At each iteration, the velocity and the position of each particle are updated according to its previous best position ($pbest_{i,j}$) and the global best position ($gbest_{i,j}$). In PSO, there are two update functions: velocity (v) and position (x) and the inertia weight (w) as given in equation (1), (2), and (3) respectively.

$$v_{i,j}(t+1) = w \cdot v_{i,j}(t) + c_1 \cdot R_1 \cdot (pbest_{i,j} - x_{i,j}(t) + c_2 \cdot R_2 \cdot (gbest_{i,j}(t) - x_{i,j}(t)) \quad (1)$$

$$x_{i,j}(t+1) = x_{i,j}(t) + v_{i,j}(t+1) \quad (2)$$

$$w(t+1) = wmax - \left(\frac{wmax - wmin}{t_{max}}\right) \cdot (t+1) \quad (3)$$

Where i is the index of particle in the swarm (i = 1,...,n); n is the population size, j is the index of position in the particle (j = 1,…,m); m is the dimension size, t represents the iteration number, $v_{i,j}$ is the velocity of the ith particle, $x_{i,j}$ is the position. Note that R_1 and R_2 are random numbers uniformly distributed between 0 and 1, c_1 and c_2 are the acceleration coefficients, and w is the positive inertia weight.

In the swarm population, there are two topologies [6]: *gbest* and *lbest* as shown in Fig.1. In the *gbest* topology, each particle is fully connected; the trajectories oscillate using a combination of an attraction to the best solution that they individually have found, and an attraction to the best solution that any particle in their neighbourhood has found. The *lbest* topology allows each individual to be influenced by some number of adjacent numbers of the population array. There are two *lbest* topologies: a von Neumann topology in which each particle is connected to its four neighbor particles and a ring topology in which where each particle is connected to its two neighbor particles.

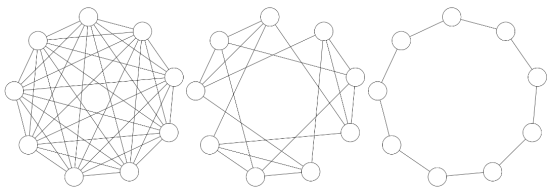


Fig.1 Particle Swarm Topologies: The *gbest* topology (left) and the *lbest* topology (center and right)

In traditional PSO, each particle in the population flies to its previous best position and the global best position [7, 8]. Although, the mechanism of this motion can result in a fast convergence rate, it suffers from the premature convergence problem, where it can be easily trapped into local minimum when solving multimodal problems. When particles move toward the direction of *gbes*t topology, the diversity among particles gradually decreases. The search is restricted to a small search space containing these similar particles. Sometime, the suboptimum are near to the global optimum and the neighborhoods of trapped individuals may contain the global optimum. It is known that the performance of PSO is related to diversify of particles, especially when attempts are made to avoid premature convergence and to escape from local optimum. Thus, in order to improve the performance of PSO, different type of *lbest* topologies such as ring lattice and Von Neumann have been proposed [9]. In this paper, we proposed another *lbest* topology using a circle to find the agent particle based on particles distance or radius.

III. Radius-Particle Swarm Optimization

A. The Algorithm

The main idea of the R-PSO is based on the *lbest* circle topology to find the agent particle within the radius of circle as shown in Fig.2. In R-PSO, a neighborhood is defined for each individual particle as the subset of particles. Each particle in the overlap radius can be in multiple groups. Once, a group is defined, we find the best particle in that group and assign to the agent particle represented by $abest_{i,j}$ as shown in Fig.3. Finally, the agent particle is the candidate for finding the optimal solution or the *gbest* position as shown in Fig.4.

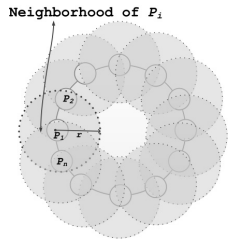


Fig. 2 The *lbest* circle topology and candidate agent particles.

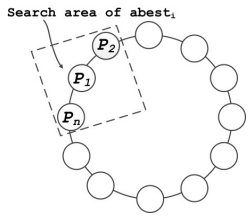


Fig.3 The local neighborhood search for the best agent particle (*abest*).

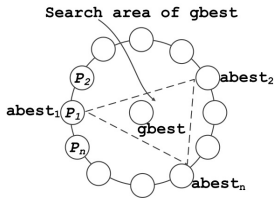


Fig.4 The global neighborhood search for the *gbest*.

Thus, the distance (d) between particles in the radius-neighborhood is computed using equation (4) and (5) respectively.

$$d(p_i, p_\emptyset) = \sqrt{\sum_{j=1}^{m} (p_{i,j} - p_{\emptyset,j})^2} \quad ; d \le 2r \qquad (4)$$

$$r = \mu \cdot vmax \qquad ; \mu \in [0.1, 0.9] \qquad (5)$$

Where d is the distance, p is the particle in the radius-neighborhood search (p = 1,..,Ø); Ø is the population size, j is the index of position in the particle (j = 1,…,m); m is the dimension size, r is the radius value, μ is the constant number μ □ [0.1,0.9], and vmax is a limit of each particle‘s velocity to a specified maximum velocity. The maximum velocity is calculated from the distance between the bounds of search space or the feasible bounds in the benchmark function. In this paper, we consider the problem of finding the global optimum using the agent particle $abest_{i,j}$ within a radius of local optimum as given in equation (6).

$$abest_{i,j} = \min_{\beta \in \emptyset} f(\beta) \qquad (6)$$

Where f is the fitness function, ß is the index of particle in the radius-neighborhood (ß = 1,…,Ø). Therefore, the position is applied to the entire velocity update equation:

$$v_{i,j}(t+1) = w \cdot v_{i,j}(t) + c_1 \cdot R_1 \cdot (pbest_{i,j} - x_{i,j}(t) + c_2 \cdot R_2 \cdot (abest_{i,j}(t) - x_{i,j}(t)) \quad (7)$$

The pseudo-code of the R-PSO algorithm is given in Fig.5.

```
for each time step t to t_max do
   for each particle i in the swarm do
      update position x_i,j(t+1) with Eq.(2) and Eq.(7)
      calculated particle's fitness f(x_i,j(t+1)
      if particle's fitness f( x_i,j(t+1) better than
         pbest x_i,j(t+1) then update pbest x_i,j(t+1)
      end if
      find agent particle
      if agent particle's fitness f(agent_i,j(t+1))better than
        gbest then update gbest
      end if
   end for
end for
```

Fig.5. R-PSO Algorithm

B. Parameter selection

The performance of R-PSO depends on the *radius* value. Therefore, experiments are carried out to find the most appropriate value of radius. The fine tuning is performed for the five well-known test functions are given in Table I and II.

TABLE I. BENCHMARK FUNCTIONS

F	Common name	Expression
f1	Sphere	$\text{Min } f(x) = \sum_{i=1}^{m} x_i^2$
f2	Schwefel 1.2	$\text{Min } f(x) = \sum_{i=1}^{m} \left(\sum_{j=1}^{i} x_j\right)^2$
f3	Generalized Rosenbrock	$\text{Min} f(x) = \sum_{i=1}^{m} 100\{(x_{i+1} - x_i^2)^2 + (x_i - 1)^2\}$
f4	Generalize Schwefel 2.6	$\text{Min} f(x) = -\sum_{i=1}^{m} x_i \sin(\sqrt{x_i})$
f5	Generalize Rastrigin	$\text{Min} f(x) = \sum_{i=1}^{m} \{x_i^2 - 10\cos(2\pi x_i) + 10\}$

TABLE II. OPTIMA AND INITIALIZATION RANGES FOR ALL FUNCTION

F	Feasible Bounds	Dimension	Optimum	Initialization Bounds
f1	$(-100,100)^m$	30	0	$(50,100)^m$
f2	$(-100,100)^m$	30	0	$(50,100)^m$
f3	$(-30,30)^m$	30	0	$(15,30)^m$
f4	$(-500,500)^m$	30	0	$(250,500)^m$
f5	$(-5.12,5.12)^m$	30	0	$(2.56,100)^m$

We calibrate our aforementioned algorithm against traditional PSO algorithm [5] and Constricted Gbest PSO Algorithm (CPSO) [10] in order to find the most appropriate radius for R-PSO. Each algorithm runs on the five benchmark functions given in Table I. The parameter w is set from the recommended value by Shi and Eberchart [11] with a linear decreasing from 0.9 to 0.4. The acceleration constants c_1, c_2 are both 1.42694, xmax and vmax are set to be equal, population size in the swarm is 60 and iteration number is 5000. For CPSO, the constriction factor $\chi = 0.72984$ and c_1, c_2 are both 2.05. A total of 30 runs for each experimental setting are conducted and average fitness of the best solution throughout the run is recorded. In the case where a gbest value is $< 10^{-8}$, it is rounded to 0.0.

The value of *Radius* may affect the performance of R-PSO. Thus, improper values of *Radius* may potentially lead the algorithm to be trapped in the local minimum. To select the better value of *Radius,* we define a scale factor μ parameter under variant μ □ [0.1, 0.9] to be analyzed. Table III present the optimum result from the benchmark functions. It has shown that the R-PSO performs best for μ is 0.4. We will employ this number in our R-PSO to measure the performance against traditional PSO and solve the RCPSP.

TABLE III. SOLUTION TO EACH RADIUS VALUE

μ	f1	f2	f3	f4	f5
0.1	0.00	2830.79	90.47	5569.33	133.63
0.2	0.00	6998.51	100.25	5355.56	167.05
0.3	0.00	536.85	3040.54	6025.58	116.49
0.4	**0.00**	**0.012**	**0.68**	**4051.17**	**18.00**
0.5	0.00	5062.98	572.89	5123.18	197.71
0.6	0.00	279.98	225.78	5891.35	31.83
0.7	0.00	9032.17	25.22	5825.92	19.91
0.8	1000.00	25058.84	3040.65	4049.57	22.88
0.9	2000.00	10405.57	25.51	4445.10	80.66

IV. RESOURCE-CONSTRAINED PROJECT SCHEDULING PROBLEM (RCPSP)

A. Problem description

During the last decades, the resource-constrained project scheduling problem is a one of the most important problems in the context of project scheduling. A common characteristic of the RCPSP is a non-pre-emptive activity. Therefore, a sequence of activity affects the duration of the project. The RCPSP is defined by set of activities and set of resources. All activities have precedence and resource constraints. The typical RCPSP is defined as follows:

1) The objective is to find the minimal makespan schedule.
2) There are N+2 activities. Let A = {1,2,3,…,n, n+2} is a set of activities which has to be processed where n is the number of activity. Activity number 1 and number N+2 refer to the start and end of schedule, respectively.
3) Each activity has precedence and resource constraints. Let P_i be the set of predecessor of activity j; the activity j cannot start to work until all of its predecessors have finished.
4) Activity 1 is the source that has no predecessors and can start immediately.
5) Let R = {1,2,3,…,n} is a set of resources required by each activity. Resources are renewable or made available at the end of the activity.

Example of precedence and resource constraints of the activity graph is given in Fig. 6.

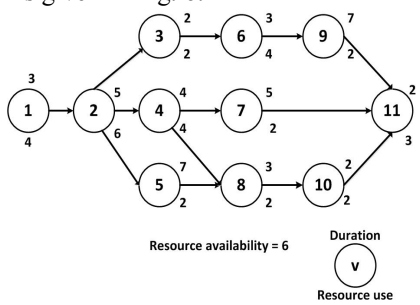


Fig. 6. The activity graph.

Fig.6 shows the activity graph with 11 activities and 6 units of available resource. Fig 7 shows two feasible schedules of this activity graph. The schedule with smaller latest finish time (LFT) is the better, hence the schedule (b) (LFT=29) performs better than the schedule (a) (LFT=31). However, the set of active schedules is usually very large and contains many schedules with relatively large delay times with poor quality in terms of makespan.

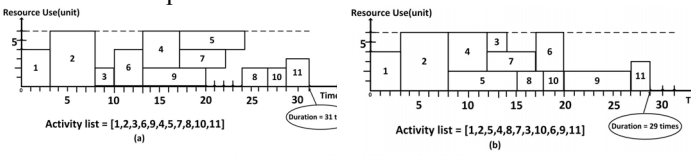


Fig. 7. Two feasible schedules.

As the computational complexity increases, Fang and Wang [12] prove that RCPSP is NP-hard. As the result, the convention exact algorithm such as branch and bound method can only solve small-size instance of the RCPSP in acceptable time. For the large-size instances, the meta-heuristic methods such as genetic algorithm (GA) [9, 13], simulated annealing (SA) [14], and tabu search (TS) [15] are more widely used to obtain the optimum solution. In addition, the advantages of the meta-heuristic [16, 17] include computational feasibility and effectiveness with acceptable time and memory requirements. However, there still exist some problems, if the solution is in a local optimum, the algorithm will not be able to search in the solution space. Hence, the algorithm plunges into the local optimum.

In meta-heuristic method, priority rule based heuristics are most popular. They employ priority rules to decide with activity to be selected from decision set in each step of constructing a solution. Thus the schedule generation scheme (SGS) is the core of most meta-heuristic procedures for the RCPSP to decode representation of a solution.

B. *Radius-Particle Swarm Optimzation for RCPSP*

1) *Framework of R-PSO for RCPSP*

This section presents the proposed R-PSO for solving the RCPSP. In general, a solution to the RCPSP is based on building a schedule from an activity. There are two form of solution: permutation and priority representation. An activity list is a precedence feasible list of all activity of the given project. Kolisch [18] and Kolisch and Hartmann [19] conclude from experimental tests that procedures based on activity list representations outperforms the other procedures. A schedule can be built using serial (S-SGS) or parallel schedule generation scheme (P-SGS). In S-SGS, activities are scheduled in the order in which they all appear in the activity list. There are schedules at the earliest start time (EST) in which the required resource become available. In P-SGS, a priority is maintained, at each point in time, precedence activity with feasible resource is scheduled, more than one activity can be assigned at a time. Since P-SGS is sometimes unable to obtain an optimum schedule [20]. In this paper we focus on the priority representation and S-SGS is adopted in our procedure.

The general framework of the R-PSO for RCPSP is illustrated in Fig.8. After generating the initial population and initaizing of the position value in each dimenion. An interative proceduce is carried out to find the optimum solution until the stop criteria is met.

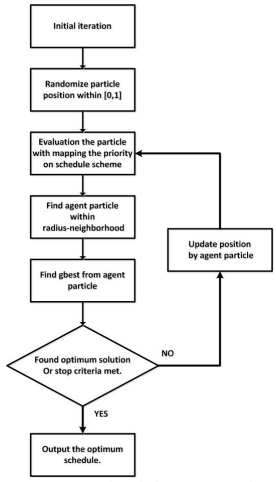


Fig.8. Framework of the R-PSO for RCPSP.

2) *Priority-based representation*

Suppose, there is a problem with 11 activities, the particle will fly in the search space with 11 dimensions. Each dimension in particle represents an activity and its corresponding value shows the random priority in the range of [0,1], the higher value represents the higher priority as shown in Fig.9.

j_0	j_1	j_2	j_3	j_4	j_5	j_6	j_7	j_8	j_9	j_{10}
A1	**A2**	**A3**	**A4**	**A5**	**A6**	**A7**	**A8**	**A9**	**A10**	**A11**
0.9	**0.85**	**0.7**	**0.8**	**0.75**	**0.65**	**0.6**	**0.55**	**0.5**	**0.45**	**0.4**

Fig.9. Representation of priority in each dimension

Where j is the index of position in the particle (j = 1,…,m);m is the dimension size, and A is the activity number; (A = 1,..,n+2);n is the number of activity.

3) *S-SGS from particle-represented solution*

To apply R-PSO for the RCPSP, the position value represents the priority of each activities (see Fig.9). To evaluate the fitness of each particle, we need to generate the SGS from the position value and find the makespan. The pseudo-code of R-PSO combines with the SGS is presented in Fig. 10.

```
Function SGS
Input particle position
Output makespan
Initiaize Seg_0 ← 1; F_0 ← 0; Dm_0 ← {∅}; t ← 0
 for k ∈ K do initialize R_k(0) ← R_k
for g = 2 to n + 2 do
    calculate Dm_g, F_g and R_k(t); k ∈ K, t ∈ F_g
    j* ← max{x_ij(t + 1) | j − 1 ∈ Dm_g}
    EST_j* ← max{F_θ | θ ∈ P_j*}
    if R_k(t) < R_j*,k ; t ∈ EST_j* + e_j* , k ∈ K then EST_j* ← EST_j* + 1
    LFT_j* ← {EST_j* + e_j*}
    Sec_g ← S_g−1 ∪ j*
    F_g ← F_g−1 ∪ LFT_j*
    R_k(t) ← R_k(t) − R_kj(t); t ∈ [LFT_j* − e_j*], k ∈ K
end for
makespan ← max_j∈n+2{F_j}
Return makespan
end function
```

Fig. 10. Pseudo-code for R-PSO combines with P-SGS.

g	Dm_g	$x_{i,j}$	j^*	Sec_g	LFT_j	F_g
1			1		4	4
2	2	0.85	2	1	8	4
3	3,4,5	0.7,0.8,0.75	4	1,2	12	4,8
4	3,5,7,8	0.7,0.75,0.6,0.55	5	1,2,4	16	4,8,12
5	3,7,8	0.7,0.6,0.55	3	1,2,4,5	14	4,8,12,16
6	6,7,8	0.65,0.6,0.55	6	1,2,4,5,3	17	4,8,12,16,14
7	7,8,9	0.6,0.55,0.5	7	1,2,4,5,3,6	20	4,8,12,16,14,17
8	8,9,11	0.55,0.5,0.4	8	1,2,4,5,3,6,7	20	4,8,12,16,14,17,20
9	9,10,11	0.5,0.45,0.4	9	1,2,4,5,3,6,7,8	24	4,8,12,16,14,17,20,20
10	10,11	0.45,0.4	10	1,2,4,5,3,6,7,8,9	22	4,8,12,16,14,17,20,20,24
11	11	0.4	11	1,2,4,5,3,6,7,8,9,10	26	4,8,12,16,14,17,20,20,24,22
				1,2,4,5,3,6,7,8,9,10,11		4,8,12,16,14,17,20,20,24,22,26

Fig. 11. Priority Schedule Generation Scheme (P-SGS).

The SGS consists of stages (g = 1,…, n+2), where n is number of active activities in the project, j* is a selected activity, P_{j*} is predecessors activity of j*, e_{j*} is a duration, EST_{j*} is the earliest start time and LFT_{j*} is the latest finish time. Let $R_k(t)$ be the remaining availability of the resource of type k at time (t) where k is the available resources (k = 1,…,K); and K is the type of resources. There are three sets of data in each g state (see Fig 11).

- Sec_g is a set of the feasible schedule.
- F_g is a set of the LFT for all activities of Sec_g .
- Dm_g is a set of selectable activities.

In each stage, an activity is selected and scheduled as early as possible according to precedence relationship and resource constraints. The SGS procedure is presented below:

a) Initialize set of the feasible schedule Sec_g that contains only activity 1 as this is the beginning of the project and Dm_g is the empty set.
b) In each stage g compute the set of selectable activities Dm_g and remaining capacity of each resource $R_k(t)$.
c) Assign a priority level to a set of selectable activities Dm_g according to the particle position in each dimension. Select activity j* with the highest priority level from the set of selectable activities Dm_g .
d) Compute earliest start time EST_{j*} in term of precedence constraint.
e) Verify the remaining resources (see Fig. 12).
f) Compute last finish time LFT_{j*} in term of duration and resource constraint.
g) Input activity j* in to the feasible schedule Sec_g.
h) Input the LFT of activity j* in to the set of F_g.
i) Update the remaining resource capacity $R_k(t)$.
j) If g is not equal N+2 go to (b).
k) The makespan is the maximum F_g.

4) Particle-updating mechanism for R-PSO

The particle position update is updated by the agent particle. First, the R-PSO are assigns the best particle in the radius-neighborhood to the agent particle. Then, the particle position are updated by the agent particle . Since, the RCPSP is the complex problems which have more than one optimum or exists on global optimum and several local optimum. The R-PSO takes advantage of group-swarm to maintain the swarm diversity and evolution by sharing information from the agent particles which effectively keep the balance between the global exploration and the local exploitation. After updating, the agent particle guides the neighbor particles to jump out of the local optimum and achieve the global best.

5) Termination

The schedule with the minimum makespan is obtainted by the return particle as the output shown as Fig. 11 and 13.

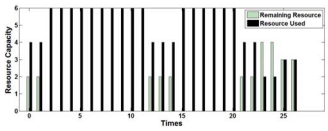


Fig. 12. Remaining resources in each period.

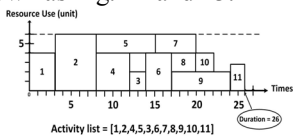


Fig. 13.The optimum schedule

V. EXPERIMENTAL SETUP

A. Dataset

The experiment used the well-known dataset PSPLIB [21] for testing. This dataset contains subset J30, J60, and J120 which have 30, 60, and 120 activities respectively. The set J30 and J60 consists of 480 instances, and J120 consists of 600 instances. Hence, there are total of 1560 instances. The PSPLIB dataset has 4 resources and each activity has 3 successors and can be downloaded from the website http://www.om-db.wi.tum.de/psplib/datasm.html.

B. Parameter configruation

The acceleration constants c_1, c_2 are both 1.42694, xmax and vmax are set to equal and in the range of [0,1]. The parameter w used a linear decreasing, which change from 0.9 to 0.4. For the R-PSO, μ is 0.4. The termination condition for each instance is the limit of the number generated solution (schedules) such as 1000 or 5000. The number of generated schedules can be calculated by Eq. (8).

$$Generated\ schedules = Populationsize\ x\ \ Iteration\ number \qquad (8)$$

The number of generated schedules is 960 (60 particles x16 iterations) for a limit of 1000 schedules and 4,980 (60 particles x 83 iterations) for a limit of 5000 schedules.

VI. RESULT AND DICUSSION

We run the R-PSO 30 times independently for each instance and set the maximum number of schedule to 1000 and 5000 as the stop criterion. The average response variable (ARV) values for J30, J60, and J120 are the average deviation values for the total instance (B) is 480, 480 and 600 instances respectively. The ARV and the optimum rate are given in Eq. (9) and Eq. (10) respectively.

$$ARV = \left(\sum_{l=1}^{B}\sum_{k=1}^{30}\frac{(OPT_{Lower\ bound} - makespan_{k,l})\ x\ 100\%}{OPT_{Lower\ bound}}\right)/(30 \cdot B) \quad (9)$$

$$Optimum\ Rate = \frac{Optimum\ Instance}{Total\ Instance\ (B)}\ x\ 100 \qquad (10)$$

Where $makespan_{k,i}$ is the makespan of the lth instance obtained by the R-PSO in the kth and $OPT_{Lower\ bound}$ is the lower bound optimum.

Table IV presents the average percentage deviation from lower bound optimum and Table V presents the dataset optimum rate. For all dataset, the R-PSO is best with both 1000 and 5000 schedule, i.e. for J30 ARV is 0.38% with 1000 schedules and 0.12% with 5000 schedules, and optimal rate is 89.79% and 94.79% respectively. Thus, the makespan

obtained by R-PSO is much close to the optimal solution than any other methods.

The comparative study also investigates the performance of our method against other meta-heuristic approaches, including genetic algorithm (GA), simulated annealing (SA) and tabu search (TS) in term of the average percentage deviation. The results show that the R-PSO is better than other methods for solving RCPSP with medium and large scales. We can see that the number of schedules has positive effect on the performance of the R-PSO. Moreover, the deviation values for all the datasets are very small, hence the R-PSO is very robust. In term of the optimum rate, our R-PSO is also outperforms the PSO and the bee swarm optimization (BSO).

TABLE IV. AVERAGE PERCENTAGE DEVIATIONS FROM THE CRITICAL PATH BASED LOWER BOUND

Algorithm	Dataset	ARV	
		1000 schedules	5000 schedules
R-PSO	J30	**0.38**	**0.12**
	J60	**12.20**	**11.83**
	J120	**34.75**	**33.60**
PSO	J30	0.41	0.33
	J60	12.99	12.54
	J120	35.05	34.02
GA[13]	J30	2.08	1.59
	J60	12.57	11.86
	J120	39.36	36.57
SA[14]	J30	0.38	0.23
	J60	12.75	11.90
	J120	42.81	37.68
TS[15]	J30	0.86	0.44
	J60	13.80	13.48
	J120	N/A	N/A

TABLE V. DATASET OPTIMUM RATE

Algorithm	Dataset	Total	Optimum Rate			
			1000 schedules		5000 schedules	
R-PSO	J30	480	**431**	**89.79**	**455**	**94.79**
	J60	480	**332**	**69.16**	**359**	**74.79**
	J120	600	**121**	**20.16**	**138**	**23.00**
PSO	J30	480	422	87.91	436	90.83
	J60	480	329	68.54	351	73.12
	J120	600	106	17.66	128	21.33
BSO[3]	J30	480	372	77.30	412	85.63
	J60	480	305	64.38	340	70.63
	J120	600	102	17.00	135	22.50

VII. CONCLUSION

We propose the R-PSO for solving the RCPSP by regrouping the particles within a given radius and determine the agent particle which is the best particle of the group. We present the process and design mechanisms to solve the problem by using serial schedule generation schemes (SGS) and map the schedule onto the particles position and calculate the fitness function. The result demonstrates that the R-PSO outperforms the traditional PSO in term of makespan and standard deviation in finding the optimal solution.

REFERENCES

[1] S. Hartmann and D. Briskorn "A survey of variants and extensions of the resource-constrained project scheduling problem," European Joural of Operation Research, vol. 207, 2010, pp. 1-14.

[2] A. Lim , H. Ma, B. Rodrigues , S.T. Tan and F. Xiao "New concepts for activity float in resource-constrained project management," Computers & Operations Research, vol. 38,2011 , pp. 917-930.

[3] K. Ziarati, R. Akbari and V. Zeighami "On the performance of bee algorithms for resource-constrained project scheduling problem," Applied Soft Computing, vol. 11, 2011, pp. 3721-3733.

[4] A.Agarwal, S. Colak and S. Erenguc "A Neurogenetic approach for the resource-constrained project scheduling problem," Computers & Operations Research, vol. 38, 2011, pp. 44-50.

[5] J. Kennedy, and R. Eberhart . "Particle swarm optimization," in IEEE international conference on neural networks, Proceedings Vol. 4, 1995 pp. 1942–1948.

[6] J. Kennedy, and R. Eberchart,"Small worlds and mega-minds: effects of neighborhood topology on particle swarm performance," in proceedings of IEEE congress on Evolutionary Computation, 1999, pp. 1391-1938,

[7] M.Gang, Z.Wei and C. Xiaolin . "A novel particle swarm optimization base on particle migration," Applied Mathematics and Computation. Vol. 218, 2012, pp. 6620–6636.

[8] J.Zhang, and X. Ding . "Multi-Swarm Self-Adaptive and Coperative Particle swarm optimization," Engineering Application and Artificial Intelligence, Vol. 24, 2011, pp. 958–967.

[9] J. Kennedy, and R. Mendes, "Population structure and particle swarm performance," in Proceeding of IEEE Congress on Evolutionary Computation, 2002, pp. 1671-1676.

[10] Y. Shi, and R.C. Eberhart, "Empirical Study of Particle Swarm Optimization," in Proceedings of the IEEE Congress on Evolutionary Computation (CEC), 1999, pp 1945-1950.

[11] M.Clerc, and J. Kennedy, "The particle swarm – explosion, stability, and convergence in a multidimension complex space," IEEE Transaction on Evolutationary Computation 6, 2002, pp. 58-73.

[12] C. Fang, and L. Wang, "An effective shuffled frog-leaping algorithm for resoure-constrained project scheduling problem," Computer & Operation Reserch, vol.39, 2012, pp. 890-901

[13] J. Alcaraz, and C. Marotoa,"A robust genetic algorithm for resource allocation in project scheduling," Annals of Operations Resource, vol 102, 2001, pp. 83-109.

[14] K. Bouleimen, and H. Lecocq,"A new efficient simulated annealing algorithm for the resource-constrainted project scheduling problem and its multiple mode version," Juropean Journal of Operational Resource, vol, 149, 2003, pp.268-281.

[15] T. Baar, P. Brucker, and S.Knust, "Tabu-search algorithms and lower bounds form the resource-constrainted project scheduling problem,"in S Voss, S Marello, I Osman, c Roucairal (Eds), Meta-heuristics: Advances and Trends in Local Search Paradigms for Optimization, Kluwer, Dordrecht, 1998, pp 1-8

[16] R. Samani," A competitive magnet-based genetic algorithm for solving the resource-constrained project scheduling problem," European Joural of Operation Research, vol.229, 2013, pp. 552-559

[17] Q. Jia, and Y. Seo," Solving resource-constrained project scheduling problems: Conceptual validation of FLP formulation and efficient permutation-based ABC computation," Computer & Operation Reserch, vol. 40, 2013, pp. 2037-2050

[18] R. Kolish."Efficient priority rules for the resource-constrained project scheduling problem," Journal of Operatations Manangement, vol 14, 1996 pp.179–192.

[19] R. Kolisch and S. Hartmann ."Heuristic algorithm for solving the resource-constrained project scheduling problem:classification and computational analysis," in J. Weglarz(Ed.), Project scheduling, recent models, algorithm and application , 1999 pp.147–178.

[20] R. Kolish."Serial and parallel the resource-constrained project scheduling methods revisted:Theory and computation," European Journal of Operational Research, vol 90, 1996 pp.320–330.

[21] R. Kolisch and A. Sprecher."PSPLIB—a project scheduling problem library," European Journal of Operational Research, vol 96, 1996 pp.205–216

Restoration of Non-Uniformly Warped Noisy Images Based on Coarse-to-Fine Optical Flow Estimation

Kalyan Kumar Halder*, Murat Tahtali, and Sreenatha G. Anavatti
School of Engineering and Information Technology
The University of New South Wales
Canberra, ACT 2600, Australia
*Email: k.halder@student.unsw.edu.au

***Abstract*—This paper proposes a high accuracy and fast image restoration approach to restore a sequence of atmospheric turbulence degraded frames of a remote object or scene. A coarse-to-fine optical flow technique is employed to estimate the dense motion fields of the frames against a reference frame. The First Register Then Average And Subtract (FRTAAS) method is used to correct the geometric distortions and restore a high quality sequence. Finally, a non-local means filter is applied to extract noise from each frame of the sequence. A performance comparison is presented between the proposed restoration method and an earlier method in terms of computational time and accuracy. The effectiveness of the proposed approach is demonstrated on both synthetic and real-life videos.**

***Keywords*—Atmospheric turbulence; image denoising; image registration; image restoration; and optical flow**

I. Introduction

For long-range imaging systems, the prevailing effects of atmospheric turbulence comprise random geometric distortions as well as non-uniform image blurring [1]–[4]. The image degradation effects arise from random inhomogeneities in the temperature distribution of the atmosphere, causing variations of refractive index along the optical transmission path which are the most prominent near to the ground [4]–[6]. Beside this, there are various particles present in the atmosphere due to pollution and natural causes such as water droplets, ice crystals, winds lifting the soil, volcanic eruptions, meteor impacts, etc. These cause significant effects in the fields of surveillance and astronomy where undisturbed image is extremely important [7], [8]. Over the last few decades, several image processing techniques have been proposed for the compensation of blurring effects. Huebner and Greco [4] present a performance comparison of four blind deconvolution methods that are applied to the restoration of turbulence degraded images. The methods are: linear Inverse Wiener Filter (IWF), non-linear Lucky-Richardson Deconvolution (LRD), Iterative Blind Deconvolution (IBD), and deconvolution using Principal Component Analysis (PCA). Among the methods, PCA based blind deconvolution provides better results, especially shorter processing time and more robustness to noise [6]. In [7], the authors propose a blur identification and image restoration method by minimizing Second-Order Central Moment (SOCM) of images. A lucky imaging system is proposed in [8] to restore astronomical images degraded by atmospheric turbulence.

When dealing with geometric distortions compensation, image registration is an important part which computes geometric deformations using the corrupted frames. There are several image registration algorithms, e.g., differential elastic image registration, B-spline based non-rigid image registration, gradient based optical flow and cross-correlation. A simple First Average Then Register (FATR) method is proposed in [9] for image restoration using a hierarchically windowed cross-correlation technique. A gradient based optical flow technique is proposed in [10] for improving the performance of the FATR method. The FATR method uses the average of a large number of individual frames as the prototype frame which results in a motion blurred image. In order to eliminate the limitations of the FATR method, the FRTAAS method [1], [2] is proposed where the prototype frame is chosen from the frames by a sharpness metric. The FRTAAS method computes the warping of each region of individual frames and determines the corrected shifts required to restore each frame. The FRTAAS, in conjunction with differential elastic registration, provides somewhat accurate and stable results though it is computationally intensive. A variant of the FRTAAS, namely FRTAASv method based on Minimum Sum of Squared Differences (MSSD) registration [11] and non-rigid registration [3] are proposed for faster implementation. Mao and Gilles [12] propose a novel approach for image restoration based on a variational model solved by Bregman Iterations and operator splitting method. Further modification is needed for the above methods in order to incorporate all other atmospheric effects such as noise and blurring.

This paper addresses the problem of how to remove the geometric distortions and noise caused by the atmospheric turbulence. We propose an improved method to recover the undistorted sequence of a remote object given only a video sequence of that object with severe distortions. We use a coarse-to-fine optical flow estimation algorithm for image registration to utilize its high accuracy and robustness property. The performance of the proposed method is compared with the MSSD registration based restoration method in terms of processing speed and accuracy.

II. Optical Flow Estimation

The optical flow method calculates the motion fields between two image frames which are taken at different times t and $t+\Delta t$. Let the reference frame and source frame be $F(\mathbf{p})$ and $F(\mathbf{p}+\mathbf{w})$, where $\mathbf{p}=(x,y,t)$ and the optical flow field $\mathbf{w}(\mathbf{p})=\big(x_s(\mathbf{p}),y_s(\mathbf{p}),1\big)$. $x_s(\mathbf{p})$ and $y_s(\mathbf{p})$ are the horizontal and vertical components of the flow field, respectively. According to the gray value constancy assumption, the gray value of a moving pixel should be consistent along the flow vector and

the flow field should be piecewise smooth [13]. This results in an objective function in the continuous spatial domain [14],

$$E(x_s, y_s) = \int \psi(|F(\mathbf{p}+\mathbf{w}) - F(\mathbf{p})|^2) + \alpha\phi(|\nabla x_s|^2 + |\nabla y_s|^2) d\mathbf{p}, \quad (1)$$

where $\psi(\cdot)$ and $\phi(\cdot)$ are robust functions, ∇ is the gradient operator and α weights the regularization. For mathematical simplicity, we use here a discrete version of (1) proposed in [14]. In this method, it is assumed that an estimate of the flow field is known as $\mathbf{w}$, and we need to estimate the best increment $d\mathbf{w}$ (i.e. dx_s and dy_s). After modifying the objective function in (1), we get

$$E(dx_s, dy_s) = \int \psi\left(|F(\mathbf{p}+\mathbf{w}+d\mathbf{w}) - F(\mathbf{p})|^2\right) + \alpha\phi\left(|\nabla(x_s+dx_s)|^2 + |\nabla(y_s+dy_s)|^2\right) d\mathbf{p}. \quad (2)$$

Using first-order Taylor expansion, the term $F(\mathbf{p}+\mathbf{w}+d\mathbf{w}) - F(\mathbf{p})$ can be written as

$$F(\mathbf{p}+\mathbf{w}+d\mathbf{w}) - F(\mathbf{p}) \approx F_z(\mathbf{p}) + F_x(\mathbf{p})dx_s + F_y(\mathbf{p})dy_s, \quad (3)$$

where $F_z(\mathbf{p}) = F(\mathbf{p}+\mathbf{w}) - F(\mathbf{p})$, $F_x(\mathbf{p}) = \frac{\partial}{\partial x}F(\mathbf{p}+\mathbf{w})$, $F_y(\mathbf{p}) = \frac{\partial}{\partial y}F(\mathbf{p}+\mathbf{w})$. The continuous function in (2) can be discretized as

$$E(dX_S, dY_S) = \sum_{\mathbf{p}} \psi\left(\left(\delta_{\mathbf{p}}^T(F_z + F_x dX_S + F_y dY_S)\right)^2\right) + \alpha\phi\Big(\left(\delta_{\mathbf{p}}^T D_x(X_S + dX_S)\right)^2 + \left(\delta_{\mathbf{p}}^T D_y(X_S + dX_S)\right)^2 + \left(\delta_{\mathbf{p}}^T D_x(Y_S + dY_S)\right)^2 + \left(\delta_{\mathbf{p}}^T D_y(Y_S + dY_S)\right)^2\Big), \quad (4)$$

where X_S, Y_S are the vectorized form of x_s, y_s; $F_x = \text{diag}(F_x)$ and $F_y = \text{diag}(F_y)$ be diagonal matrices where the diagonals are the frames F_x and F_y; D_x and D_y represent the matrix corresponding to $x-$ and $y-$derivative filters; $\delta_{\mathbf{p}}$ is the column vector that has only one nonzero (one) value at location $\mathbf{p}$, e.g., $\delta_{\mathbf{p}}F_x = F_x(\mathbf{p})$. Let

$$f_{\mathbf{p}} = \left(\delta_{\mathbf{p}}^T(F_z + F_x dX_S + F_y dY_S)\right)^2, \text{ and}$$
$$g_{\mathbf{p}} = \left(\delta_{\mathbf{p}}^T D_x(X_S + dX_S)\right)^2 + \left(\delta_{\mathbf{p}}^T D_y(X_S + dX_S)\right)^2 + \left(\delta_{\mathbf{p}}^T D_x(Y_S + dY_S)\right)^2 + \left(\delta_{\mathbf{p}}^T D_y(Y_S + dY_S)\right)^2. \quad (5)$$

We can rewrite (4) as

$$E(dX_S, dY_S) = \sum_{\mathbf{p}} \psi(f_{\mathbf{p}}) + \alpha\phi(g_{\mathbf{p}}). \quad (6)$$

We use the Iterative Reweighted Least Square (IRLS) to find dX_S, dY_S so that the gradient $\left[\frac{\partial E}{\partial dX_S}; \frac{\partial E}{\partial dY_S}\right] = 0$. Deriving $\frac{\partial E}{\partial dX_S}$ and $\frac{\partial E}{\partial dY_S}$ using (6) and setting those equal to zero we get the following equations:

$$\begin{bmatrix} \Psi' F_x^2 + \alpha L & \Psi' F_x F_y \\ \Psi' F_x F_y & \Psi' F_y^2 + \alpha L \end{bmatrix} \begin{bmatrix} dX_S \\ dY_S \end{bmatrix} = -\begin{bmatrix} \Psi' F_x F_z + \alpha L X_S \\ \Psi' F_y F_z + \alpha L Y_S \end{bmatrix}, \quad (7)$$

where L is a generalized Laplacian filter defined as $L = D_x^T \Phi' D_x + D_y^T \Phi' D_y$, $\Psi' = \text{diag}\big(\psi'(f_{\mathbf{p}})\big)$, $\Phi' = \text{diag}\big(\phi'(g_{\mathbf{p}})\big)$.

The equations in (7) contain nonlinear functions and are solved by the following fixed-point iterations [14]:

Step 1: Set $dX_S = 0$ and $dY_S = 0$.
Step 2: Calculate the "weight" Ψ' and Φ' based on the current estimate dX_S and dY_S.
Step 3: Solve the equations in (7).
Step 4: If dX_S and dY_S do not converge go to step 2, otherwise, stop.

Using few outer and inner fixed-point iterations, we compute the flow fields through a coarse-to-fine refining scheme on a dense Gaussian pyramid.

III. Image Restoration Approach

Let an atmospheric turbulence degraded video consists of N warped and noisy frames. In FRTAAS approach the sharpest frame is considered as the reference frame. The sharpness metric employed to select the sharpest frame is [15]:

$$S = -\int F(x,y) \ln\left[F(x,y)\right] \mathrm{d}x\mathrm{d}y, \quad (8)$$

where S is the sharpness value of each individual frame F. The frame is deemed "sharper" for more negative values of S. Once the reference frame is selected, the flow fields $x_s(x,y,t)$ and $y_s(x,y,t)$ are estimated for all the frames through the backward mapping optical flow technique. By using the flow fields, the centroids, C_x and C_y, which are used to calculate the pixel locations of the original frame, are calculated by averaging:

$$C_x(x,y) = \frac{1}{N}\sum_{t=1}^{N} x_s(x,y,t)$$
$$C_y(x,y) = \frac{1}{N}\sum_{t=1}^{N} y_s(x,y,t). \quad (9)$$

The inverse of C_x and C_y are then calculated as

$$C_x^{-1}(x,y) \cong -C_x(x - C_x(x,y,t), y - C_y(x,y,t))$$
$$C_y^{-1}(x,y) \cong -C_y(x - C_x(x,y,t), y - C_y(x,y,t)). \quad (10)$$

The corrected flow fields X_s and Y_s for each warped frame in the sequence are then

$$X_s(x,y,t) = C_x^{-1}(x,y) + x_s\left(x + C_x^{-1}(x,y), y + C_y^{-1}(x,y), t\right)$$
$$Y_s(x,y,t) = C_y^{-1}(x,y) + y_s\left(x + C_x^{-1}(x,y), y + C_y^{-1}(x,y), t\right). \quad (11)$$

Using these corrected flow fields, one is therefore able to obtain:

$$R(x,y,t) = F\left(x + X_s(x,y,t), y + Y_s(x,y,t), t\right), \quad (12)$$

where R is the dewarped version of the original warped frame F.

The dewarped frames still contain several unstructured and random noise. A simple noise level estimation method [16] is employed for estimating the variance of Gaussian noise for each individual frame. The method employs a zero mean operator, which is almost insensitive to frame structure. It simply requires the use of a 3×3 mask followed by a summation over the frame or a local neighborhood. A non-local means filter [17], with the estimated noise variance, is applied to extract noise from the sequence. We fix a search window of 5×5 pixels and a similarity square neighborhood of 3×3 pixels.

IV. Simulation Experiments

The proposed method was implemented in MATLAB and tested on an Intel Core i7-2600 CPU 3.40GHz machine with 8GB RAM. At first, a video sequence of 80 warped frames was generated using the Lena (512×512 pixels) test image to compare the proposed method with the MSSD based FRTAAS method [11]. Gaussian noise was added to the Lena sequence resulting in Signal-to-Noise Ratio (SNR) of 26.13dB. Later on, the method was applied on real-world videos. The optical flow estimation algorithms were written in C MEX code which reduces the processing time significantly than their MATLAB implementation.

A. Performance Analysis and Comparison

For fair comparison we employed the same image denoising algorithm for both methods. In order to register a pair of Lena frames it requires 0.92s and 0.31s for the MSSD and coarse-to-fine flow algorithms, respectively. Therefore, the proposed optical flow algorithm is almost 3 times faster than earlier MSSD algorithm on the same machine. Fig. 1(a) and (b) show the ground truth Lena frame and the reference frame, respectively. The restored versions of the reference frame using the MSSD based method and the proposed method are shown in Fig. 1(c) and (d), respectively. Fig. 1(e) and (f) present the difference between the ground truth and restored reference frame for the two methods. It can be seen that the proposed method provides lower residuals than the MSSD based method. The intensity Mean Squared Error (MSE) values of the restored reference frame are 89.44 and 58.52 for the MSSD based method and the proposed method, respectively. Fig. 2 shows a frame-wise MSE comparison for the Lena sequence. It is apparent that the proposed method yields lower MSE for all the restored frames.

Image denoising technique cannot completely remove all noise. Therefore, the performance of the methods is also compared by their restoration capability under noisy conditions. Fig. 3 shows a plot of the average MSE of the restored Lena frames for a range of SNR. The performance of the restoration degrades with increasing noise level. It is clearly evident that the proposed method works better than the MSSD based FRTAAS method under the presence of noise.

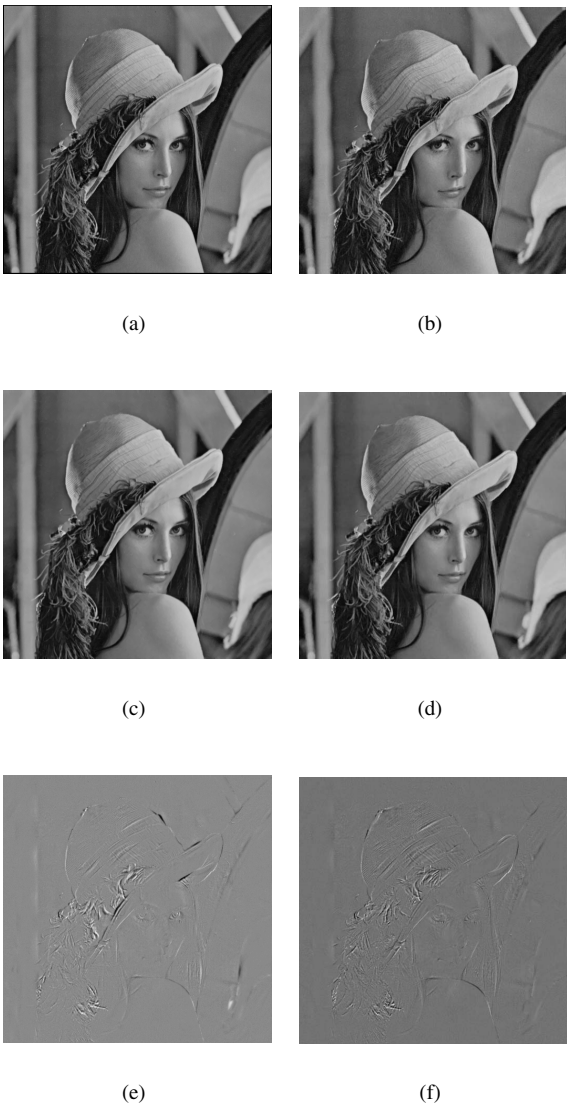


Fig. 1. (a) Ground truth frame, (b) warped reference frame, (c) restored reference frame using MSSD based FRTAAS method, (d) restored reference frame using proposed method, (e) intensity difference between (a) and (c), and (f) intensity difference between (a) and (d).

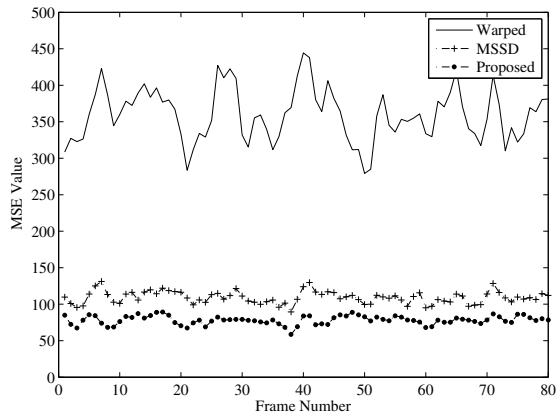


Fig. 2. MSE plot of the warped and restored frames using MSSD and coarse-to-fine flow based FRTAAS method for Lena sequence.

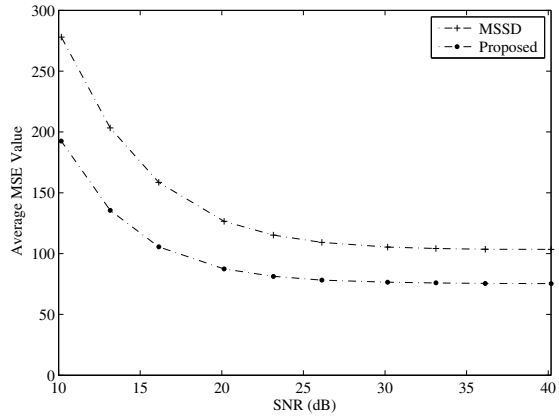


Fig. 3. Average intensity MSE of all restored frames versus SNR for Lena sequence.

B. Application on Real Data

After obtaining good restoration performance from the proposed method on synthetically warped and noisy video sequence we applied it on real-world surveillance video sequences. Fig. 4 shows the mean of all turbulent frames, and restored versions of the reference frame using the two methods for three real-world datasets [3], [18], [19]. In the case of the real video, since no ground truth image is available, the restored frames cannot be compared with the original. An objective comparison is made using the simple sharpness metric in (8) and is presented in Table I. It can be seen that the simple mean of the turbulent frames suffers from blurring whereas the restored reference frame using the proposed method is sharper than for MSSD based method.

TABLE I. COMPARISON OF SHARPNESS FOR REAL-WORLD DATASETS

Sequence	Sharpness value		
	Turbulent	MSSD	Proposed
Hill-house (512×512)	-5.7319e7	-5.7920e7	-5.8175e7
Tower (460×560)	-1.5611e8	-1.5617e8	-1.5619e8
Building-site (594×876)	-2.8369e8	-2.8381e8	-2.8403e8

(a) Hill-house

(b) Tower

(c) Building-site

Fig. 4. Mean of all turbulent frames (left), restored reference frame using MSSD based method (middle), and restored reference frame using proposed method (right).

V. CONCLUSION

We have successfully implemented a multi-frame image restoration method using a coarse-to-fine optical flow estimation technique in order to compensate for geometric distortions under noisy conditions. Results obtained from both simulated and real data have been presented. It has been demonstrated that using our proposed method, a faster restoration of the sequence is possible with improved registration accuracy. In the future, we will consider other atmospheric effects such as blurring in order to overcome the complete turbulence problem.

REFERENCES

[1] M. Tahtali, D. Fraser, and A. J. Lambert, "Restoration of non-uniformly warped images using a typical frame as prototype," in *Proc. TENCON*, 2005, pp. 1–6.

[2] M. Tahtali, A. J. Lambert, and D. Fraser, "Restoration of nonuniformly warped images using accurate frame by frame shiftmap accumulation," in *Proc. SPIE 6316*, 2006.

[3] K. K. Halder, M. Tahtali, and S. G. Anavatti, "A fast restoration method for atmospheric turbulence degraded images using non-rigid image registration," in *Proc. ICACCI*, 2013, pp. 394–399.

[4] C. S. Huebner and M. Greco, "Blind deconvolution algorithms for the restoration of atmospherically degraded imagery: a comparative analysis," in *Proc. SPIE 7108*, 2008.

[5] X. Zhu and P. Milanfar, "Image reconstruction from videos distorted by atmospheric turbulence," in *Proc. SPIE 7543*, 2010.

[6] D. Li, R. M. Mersereau, and S. Simske, "Atmospheric turbulence-degraded image restoration using principal components analysis," *IEEE Geoscience and Remote Sensing Letters*, vol. 4, no. 3, pp. 340–344, 2007.

[7] L. Yan, M. Jin, H. Fang, H. Liu, and T. Zhang, "Atmospheric-turbulence-degraded astronomical image restoration by minimizing second-order central moment," *IEEE Geoscience and Remote Sensing Letters*, vol. 9, no. 4, pp. 672–676, 2012.

[8] S. Zhang, Y. Wu, J. Zhao, and J. Wang, "Astronomical image restoration through atmosphere turbulence by lucky imaging," in *Proc. SPIE 8009*, 2011.

[9] D. Fraser, G. Thorpe, and A. Lambert, "Atmospheric turbulence visualization with wide-area motion-blur restoration," *J. Opt. Soc. Am. A*, vol. 16, no. 7, pp. 1751–1758, 1999.

[10] D. Clyde, I. Scott-Fleming, D. Fraser, and A. Lambert, "Application of optical flow techniques in the restoration of non-uniformly warped images," in *Proc. DICTA*, 2002, pp. 195–200.

[11] M. Tahtali, A. J. Lambert, and D. Fraser, "Graphics processing unit restoration of non-uniformly warped images using a typical frame as prototype," in *Proc. SPIE 7800*, 2010.

[12] Y. Mao and J. Gilles, "Non rigid geometric distortions correction-application to atmospheric turbulence stabilization," *Inverse Problems and Imaging*, vol. 6, no. 3, pp. 531–546, 2012.

[13] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert, "High accuracy optical flow estimation based on a theory for warping," in *Proc. ECCV*, 2004.

[14] C. Liu, *Beyond pixels: exploring new representations and applications for motion analysis*, Ph.D. thesis, Massachusetts Institute of Technology, 2009.

[15] R. K. Tyson, *Introduction to adaptive optics*, SPIE press, 2000.

[16] J. Immerkær, "Fast noise variance estimation," *Computer Vision and Image Understanding*, vol. 64, no. 2, pp. 300–302, 1996.

[17] A. Buades, B. Coll, and J. M. Morel, "A non-local algorithm for image denoising," in *Proc. CVPR*, 2005, pp. 60–65.

[18] J. P. Delport, "Scintillation mitigation for long-range surveillance video," in *Proc. Science Real and Relavent Conference*, 2010.

[19] J. de Villiers, F. Wilson, and F. Nicolls, "A study of the sensitivity of long-range passive ranging techniques to atmospheric scintillation," in *Proc. SPIE 8399*, 2012.

Modified CMA Based Blind Equalization and Carrier-Phase Recovery in PDM-QPSK Coherent Optical Receivers

Md. Saifuddin Faruk
Dept. of Electrical and Electronic Engineering
Dhaka University of Engineering and Technology,
Gazipur, Bangladesh
faruk@duet.ac.bd

***Abstract*—Digital coherent optical receivers enable equalization of fiber transmission impairments and carrier-phase estimation in digital domain. In this paper, a novel equalization and carrier-phase recovery algorithm, both based on modified constant-modulus algorithm (CMA), is proposed for polarization-division multiplexed quadrature phase-shift keying (PDM-QPSK) systems. The proposed scheme is validated with 40-Gbits/s PDM-QPSK transmission experiments.**

***Index Terms*—Optical fiber communication, equalization, carrier-phase recovery, digital signal processing.**

I. Introduction

Digital coherent optical receiver is a promising candidate for the receiver of next generation optical networks as it enables demodulation of higher-order complex modulation formats and equalization of fiber transmission impairments such as chromatic dispersion (CD), polarization-mode dispersion (PMD) etc. in digital domain. It also allows electronic carrier-phase estimation and thus achieving optical phase-lock loop (PLL) free coherent receivers [1].

In a blind PDM-QPSK receiver, equalization is generally done using four finite-impulse-response (FIR) filters connected in two-by-two butterfly configuration and adapted by constant-modulus algorithm (CMA) [2] whereas carrier-phase recovery is done using M-th power algorithm [3]. However, the CMA suffers from the drawback of slow convergence and M-th algorithm suffers from higher computational complexity due to involvement of several nonlinear operations such as taking M-th power, taking argument, phase unwrapping functions etc. To avoid such difficulties, a modified CMA based blind equalization and carrier-phase recovery algorithm is proposed in [4]. However, in [4] only single-polarization transmission case is consider.

In this paper, in extension of work in [4], a novel modified CMA based equalization and carrier recovery scheme is proposed which is applicable for dual-polarization QPSK transmission systems. Such equalizer also performs polarization demultiplexing. The singularity problem [5-6] in the blind polarization demultiplexing using the proposed algorithm is investigated and a technique for avoidance of such issue is demonstrated. The performance of the proposed scheme is verified with 40-Gbits/s PDM-QPSK transmission experiments and compared with that from the conventional scheme.

In the rest of the paper, superscripts $(\bullet)^T$ and $(\bullet)^*$ are transpose and conjugate operators, respectively, and subscripts $(\bullet)_R$ and $(\bullet)_I$ of a variable denote real and imaginary parts of the variable.

II. Proposed Scheme

The conventional blind equalization and carrier-phase estimation block diagram for PDM-QPSK transmission system is shown in Fig.1. Four butterfly connected multi-tap FIR filters adapted by CMA are used for linear impairments equalization. Such a circuit also performs polarization demultiplexing. After that M-th power carrier-phase estimation is done on both polarization channel.

On the other hand, the proposed scheme is shown in Fig. 2 where a multi-tap filter for equalization is followed by single-tap phase rotator for carrier-phase estimation (CPE), both adapted by the modified CMA.

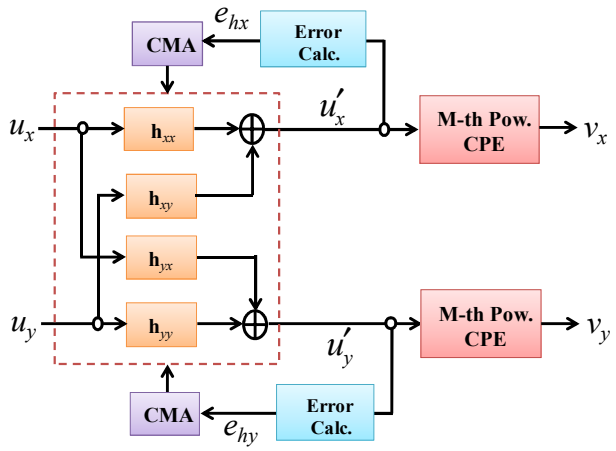


Fig.1: Block diagram of conventional blind equalization and carrier-recovery scheme for PDM-QPSK systems.

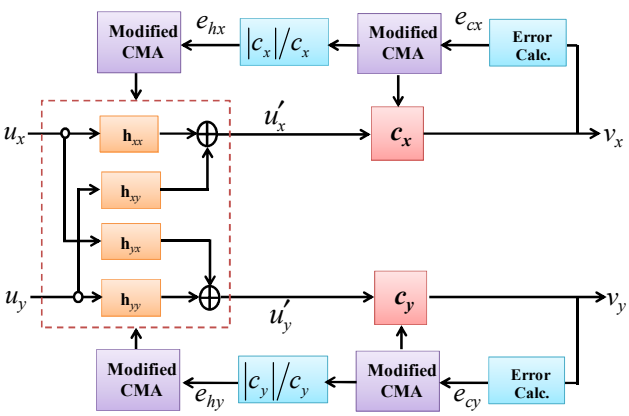


Fig.2: Block diagram of proposed blind equalization and carrier-recovery scheme in PDM-QPSK coherent optical receivers.

Consider $\mathbf{h}_{pq}(n)$ be a tap-coefficient column vector of the equalizer, where, $pq = xx, xy, yx$ or yy; $c_{x,y}(n)$ is tap coefficient of the phase rotator, and $\mathbf{u}_{xy}(n)$ a column vector of the FIR-filter input, where n is the sampling index and subscript x and y are of a variable is for x- and y-polarization channel. Then, outputs of the equalizer are given as

$$u'_x(n) = \mathbf{h}_{xx}(n)^T \mathbf{u}_x(n) + \mathbf{h}_{xy}(n)^T \mathbf{u}_y(n), \quad (1)$$

$$u'_y(n) = \mathbf{h}_{yx}(n)^T \mathbf{u}_x(n) + \mathbf{h}_{yy}(n)^T \mathbf{u}_y(n). \quad (2)$$

The output of the phase rotator is

$$v_x(n) = c_x(n) u'_x(n), \quad (3)$$

$$v_y(n) = c_y(n) u'_y(n), \quad (4)$$

The tap coefficient is updated by using the stochastic gradient algorithm as

$$c_x(n+1) = c_x(n) + \mu_c e_{cx}(n) u'^*_x(n), \quad (5)$$

$$c_y(n+1) = c_y(n) + \mu_c e_{cy}(n) u'^*_y(n), \quad (6)$$

where μ_c is the step-size parameter and $e_{cx,cy}(n)$ is the error signal. The error signal in the modified CMA [7] is computed from real and imaginary parts of the output signal as

$$e_{cx,xy}(n) = \left\{ v_{xR,yR}(n)\left(0.5 - \left|v_{xR,yR}(n)\right|^2\right)\right\} + j\left\{ v_{xI,yI}(n)\left(0.5 - \left|v_{xI,yI}(n)\right|^2\right)\right\}, \quad (7)$$

unlike the conventional CMA where error signal can be written as

$$e_{cx,cy}(n) = \left\{ v_{x,y}(n)\left(1 - \left|v_{x,y}(n)\right|^2\right)\right\}, \quad (8)$$

The modified CMA moves the real part of the output on lines of $x = \pm\sqrt{1/2}$ and the imaginary part on lines of $y = \pm\sqrt{1/2}$. Thus, it can control the amplitude and phase of the output to proper values, and carrier phase recovery is accomplished. Next, tap-coefficients of the equalizers are updated as

$$\mathbf{h}_{xx}(n+1) = \mathbf{h}_{xx}(n) + \mu_h e_{hx}(n) \mathbf{u}_x^*(n), \quad (9)$$

$$\mathbf{h}_{xy}(n+1) = \mathbf{h}_{xy}(n) + \mu_h e_{hx}(n) \mathbf{u}_y^*(n), \quad (10)$$

$$\mathbf{h}_{yx}(n+1) = \mathbf{h}_{yx}(n) + \mu_h e_{hy}(n) \mathbf{u}_x^*(n), \quad (11)$$

$$\mathbf{h}_{yy}(n+1) = \mathbf{h}_{yy}(n) + \mu_h e_{hy}(n) \mathbf{u}_y^*(n), \quad (12)$$

where μ_h is the step-size parameter and the error signal is computed by multiplying $e_{cx,cy}(n)$ with the estimated phase as

$$e_{hx,hy}(n) = e_{cx,cy}(n)\left\{\left|c_{x,y}(n)\right| / c_{x,y}(n)\right\}. \quad (13)$$

Such an approach eliminates the phase noise involved in $\mathbf{u}_{x,y}^*(n)$ of Eqs. (9-12), leading to phase-noise insensitive adaptation of the multi-tap filter with the modified CMA.

III. Avoidance of Singularity Problem

Similar to the CMA, the proposed modified CMA suffers from the singularity problem, which means both outputs converge to the same polarization tributary, thus disenabling proper polarization demultiplexing. To avoid the singularity problem, a similar technique presented in [8] is used in this work. First, the tap coefficients of $\mathbf{h}_{xx}$ and $\mathbf{h}_{xy}$ are updated using Eqs. [9] and [10] applied to output v_x. After the convergence of v_x, the values of $\mathbf{h}_{yx}$ and $\mathbf{h}_{yy}$ are set as $\mathbf{h}_{yx} = -TR[\mathbf{h}_{xy}]^*$ and $\mathbf{h}_{yy} = TR[\mathbf{h}_{xx}]^*$, where TR means time-reversed. The new equalizer matrix is then used as the initial condition for the subsequent continuous operation of all four filters using Eqs. (9)-(12). Imposing such constraint in the tap value initialization ensure to maintain orthogonality of two polarization channels and thus singularity free operation of the equalizer.

IV. Experimental Results

To prove the effectiveness of the proposed scheme, 40-Gbits/s PDM-QPSK experiments are conducted as shown in Fig.3.

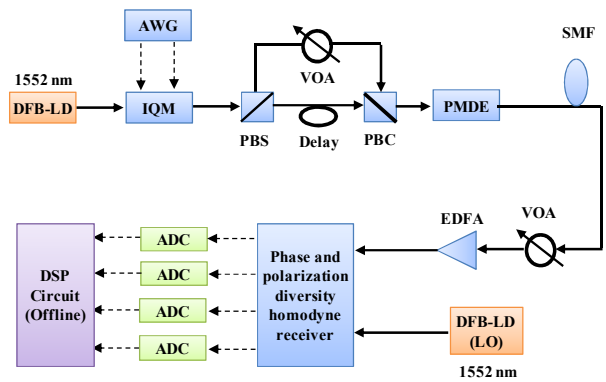


Fig.3: Experimental setup for verify the proposed scheme.

The transmitter laser as well as local oscillator (LO) are a distributed-feedback lasers (DFB-LDs) having a center wavelength of 1552 nm and a 3-dB linewidth of 150 kHz. A 20-Gbit/s NRZ-QPSK signal is generated using a $LiNbO_3$ optical IQ modulator (IQM) from two streams of precoded

data from an arbitrary waveform generator (AWG) with 2^9-1 pseudo-random binary sequences (PRBS). A 40-Gbit/s dual-polarization signal is then produced in the split-delay-combine manner by using a polarization beam splitter (PBS) and a polarization beam combiner (PBC). The delay length is set long enough so that the signals in two polarization tributaries are uncorrelated. The PDL is generated by attenuating one polarization signal using a variable optical attenuator (VOA). Then, the signal is passed through a PMD emulator (PMDE) and a standard single-mode fiber (SMF) which accumulates the CD. After the transmitted signal is pre-amplified by an erbium-doped fiber amplifier (EDFA), it is incident on a phase and polarization diverse coherent optical receiver. The outputs of the receiver are sampled and digitized at 20 Gsample/s with analog-to-digital converters (ADCs), and stored for offline digital signal processing (DSP).

In the DSP, after being kept 2 samples/symbol, polarization demultiplexing, equalization and carrier-phase recovery is done using either the conventional approach as shown in Fig.1 or by the proposed scheme as shown in Fig.2. Finally, the symbols were decoded and bit error rate (BER) is estimated as performance measure.

Figure 4 shows the BER characteristics as a function of received power. The channel impairments are set as CD=1600 ps/nm (for 100-km fiber), mean PMD of 20 ps and PDL of -3 dB. As shown in Fig.4, the proposed scheme has the performance similar to that of the conventional approach; however, it has lower computational cost. The sensitivity difference between two polarization tributaries stems from the effect of PDL.

As shown in Fig. 5, the rotating constellation at the output of the multi-tap equalizer confirms that it operates phase-insensitively. On the other hand, the clear QPSK constellation at the output of the phase rotator ensures correct phase-noise estimation.

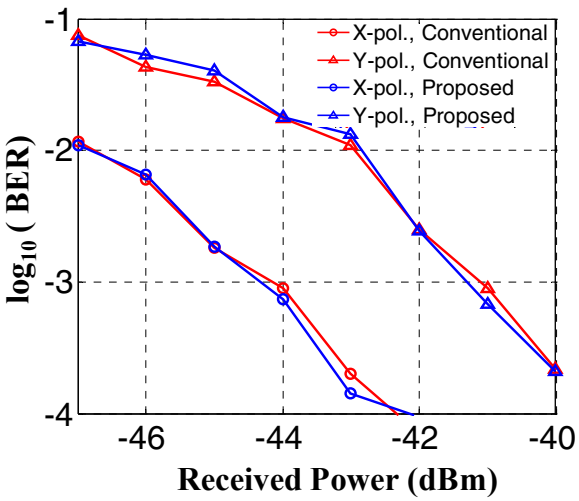


Fig.4 : BER characteristics as a function of received power for both *x*- and *y*-polarization channel.

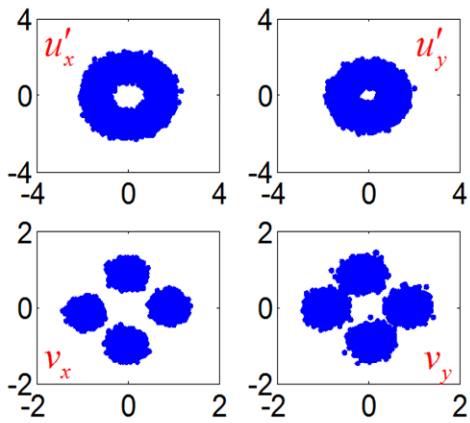


Fig.5 : Constellation map for both x- and y-polarization channel at the output of equalizer and carrier phase recovery circuit at received power of -40 dBm.

TABLE I: PERCENTAGE OF PROBABILITY OF SINGULARITY OCCURRENCE FOR DIFFERENT VALUES OF PDL

PDL	0 dB	-1 dB	-3 dB
Standard modified CMA	5%	8%	22%
Constraint modified CMA	0%	0%	0%

The probability of singularity-problem occurance depends on PDL of the system. We measure the singularity occurrence probability for CD=0 (back-to-back experiment), PMD=0 and different PDL values. The received power is set to -40 dB. As shown in Tab. I, for the standard modified CMA the singularity occurrence probability increases with increasing PDL; however, putting constraint in initial values of tap coefficients as described in Sec.3 completely avoid any singularity even in high PDL case.

V. CONCLUSION

A modified-CMA-based equalization and carrier-phase-recovery scheme is introduced for PDM-QPSK coherent optical receivers. With 40-Gbits/s PDM-QPSK transmission experiments, we have verified that the proposed scheme has the same BER performance compare to the conventional scheme of CMA based equalization and M-th power carrier-phase recovery. However, our proposed scheme has the benefit of having less computational complexity compare to the conventional approach.

REFERENCES

[1] K. Kikuchi, "Digital coherent optical communication systems: fundamental and future prospect," IEICE Electronics Express, vol. 20, pp. 1642-1662, 2011.

[2] Seb J. Savory, "Digital filters for coherent optical receivers," *Opt. Express,* vol. 16, pp. 804-817, 2008.

[3] K. Kikuchi, "Phase-diversity homodyne detection of multilevel optical modulation of with digital carrier phase estimation," *IEEE J. Sel. Topics Quantum Electron.*, vol. 12, pp. 563-570, July/August 2006.

[4] Md. S. Faruk and K. Kikuchi, "Blind equalization and carrier-phase recovery in QPSK coherent optical receivers based on modified constant-modulus algorithm," in *Proc. Optoelectronics and Communications Conference 2013*, ThR3-7, Kyoto Japan, Jun 30-Jul.4, 2013.

[5] K. Kikuchi, "Polarization-demultiplexing algorithm in digital coherent receiver," *IEEE LEOS Summer Topical Meetings2008,* MC2.2, Acapulco, Mexico, July 21-23, 2008.

[6] Md. S. Faruk, Y. Mori, C. Zhang, and K. Kikuchi, "Proper polarization demultiplexing in coherent optical receiver using constant modulus algorithm with training mode, " in Proc. *of Optoelectronics and Communications Conference 2010*, 9B3-3, Sapporo, Japan, Jul.5-9, 2010.

[7] K. N. Oh and Y. O. Chin, "Modified constant modulus algorithm: Blind equalization and carrier phase recovery algorithm," Proc. ICC, vol. 1, pp. 498 -502, 1995.

[8] L. Liu, Z. Tao, W. Yan, S. Oda, T. Hoshida, J. C. Rasmussen, "Initial tap setup of constant modulus algorithm for polarization de-multiplexing in optical coherent receivers," *Optical Fiber Communication Conference 2009,*OMT2, Los Angeles, USA, Mar. 22-26, 2009.

A Gaming Approach in Physical Therapy for Facial Nerve Paralysis Patient

Tamanna Tasneem, Atanu Shome, SK Alamgir Hossain
Computer Science and Engineering Discipline, Khulna University
Khulna, Bangladesh
{(tamannacse09, atanu.cse.ku}@gmail.com, alamgir@cseku.ac.bd

***Abstract*—This work is based on developing a system for the Bell's palsy, a type of facial nerve paralysis, patients to help them in passive exercise, which is one of the major ways to recovery. This exercise method will be provided in the form of computer game so the user will be able to perform exercise without the hardship that comes with active exercise. It will also show how much this system helps the patients and how its performance is analyzed by the users' i. e. the patients. We will also show that how passive exercise is preferable to current active one. As per initial experimentation, we found out that the proposed gaming approach for Bell's palsy recovery is useful to the patient.**

***Index Terms*—Bell's palsy, facial nerve paralysis, passive exercise, computer game**

I. Introduction

Computer science is spreading in each and every field that one can think of now-a-days. And now it's bonding with medical field as some robust composition in terms of both detection and cure of diseases. Human Computer Interaction (HCI) made it possible to shorten the gap between human and computer methods by which new possibilities and ways are opening for new treatment methodologies [1] [2] [3]. Dropping of cost, modern technical advancement, powerful tools made it possible to have larger impact on health care field which lead us to think of a way to cure Bell's palsy disease [4] .

Bell's palsy is a form of facial paralysis resulting from a dysfunction of the cranial nerve (the facial nerve) causing an inability to control facial muscles on the affected side [5] [6]. Facial nerve controls the muscles of facial expression, and functions in the conveyance of taste sensations from the anterior two-thirds of the tongue and oral cavity.

When a patient attacked by Bell's palsy disease it damaged the facial nerve [7], and facial nerve generally creates a trauma on one side of the face. Electrochemical signals sent from the brain travel along facial nerve. If the facial nerve is disrupted, no signals can get pass through these muscles, and then half side of the face gets affected by experiencing muscles weakness or paralysis, known as Bell's palsy. Some effects of Bell's palsy causes disruption in eye controlling, smiling, eating, interference in facial expressions, and sometimes speech may also blurred. Tearing of eye, interference in hearing also could be seen in Bell's palsy.

Bell's palsy affected one side of the face and it can cause closing of an eye for a certain period. Exercise is required to save the eye from severe or permanent damage. Most of the cases without any treatment Bell's palsy may recover with exercise. So doctors advise to do eye and mouth exercise to close the eye and move the mouth. But due to muscles weakness patient have trouble to control the eyes and mouth by himself. So passive exercise by the patient is effective than the active exercise. So we propose gaming technique which will help a patient to exercise in a manner that will help to exercise the eye muscles and facial nerves which will result in better retrieval from bad effects. In our proposed method patient concentrate on the game rather than exercise which is less laborious than actively open or close the eye.

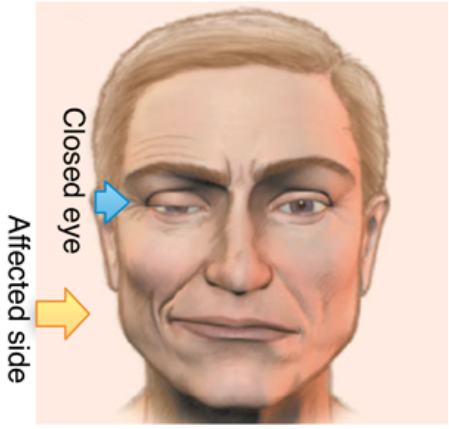


Fig. 1: A human face illustrating Bell's palsy.

Our contribution in this paper is two-fold. First, we present an algorithm by which bell's palsy patient have the opportunity to exercise with less trouble. Second, we presented the experimental result based on our developed prototype system.

The remainder of this paper is organized as the following. At first, in Section II we present some recognized work that are related to this paper. Then in Section III we illustrate the various components of our proposed technique. Further, in Section IV we present the implementation details of our prototype. The experimental results and the performance of our system is discussed in Section V. At the end we provide conclusion of the paper in Section VI and state some possible future work directions.

II. Related work

Bell's palsy is not a new disease among the physicians. At first Sir Charles Bell [8], a Scottish surgeon, in 19th century

recognized the disease. From onward various researches try to find out the treatment for this disease. The most commonly used treatment tries to find out from the patient's history and neurological examination. If weakness is central, brain magnetic resonance imaging (MRI) is required to evaluate the patient for ischemia and for infectious and inflammatory diseases. If facial weakness is peripheral, no apparent cause will be found in most instances and no tests are immediately indicated [1].

The Bell's palsy recovery and characteristics are mostly highlighted by Olsen [9]. He presented a statistical result based on 50 Bell's palsy patients. He showed facts about signs, symptoms and causes of Bell's palsy. Bell's palsy may occur one single time or may come every month or year. The Bell's palsy which comes recurrently calls recurrent Bell's palsy. It is still unknown why recurrent Bell's palsy happens in a patient. To identify nature of recurrent Bell's palsy Pitts et. al. [10] presented a statistical report based on 140 real patients. From his study he observed that in 7.1% cases had recurrence in Bell's palsy.

Apart from antiviral treatment, physiotherapy can be beneficial to some individuals with Bell's palsy as it helps to maintain muscle tone of the affected facial muscles and stimulate the facial nerve [6]. It is important that muscle re-education exercises and soft tissue techniques be implemented prior to recovery in order to help prevent permanent contractures of the paralyzed facial muscles [6]. To reduce pain, heat can be applied to the affected side of the face [11]. Physiotherapy may consist of electrical stimulation or facial exercises, or both. The facial exercises can be performed at home whilst the electrical stimulation is done by the physiotherapist at their clinic. The facial exercise involves doing different muscle movements to work each group of the facial muscles.

The computer based physiotherapy is a relatively new idea. In 2010 Alamri et. al. [2] proposes a novel approach based on augmented-reality (AR) technologies that can increase a stroke-patient's involvement in the rehabilitation process. The approach takes advantage of virtual-reality technologies and provides natural-force interaction with the daily environment by adopting a tangible-object concept.

The approach takes advantage of virtual-reality technologies and provides natural-force interaction with the daily environment by adopting a tangible-object concept. Similar work has been done before on rehabilitation of the hand of post-stroke patients in 2006 [12].

III. Proposed Framework

In this section we describe the general architecture of the proposed system. The architecture is depicted in Figure 2 as a block diagram. The Bell's palsy patient face is captured by a video feed or a standalone camera. An IR (Infra-Red) sensor is used to measure the distance between the camera and the patient face. In our system the patient have to sit between 12cm-16cm feet distance. The IR sensor measures the distance and suggests the patient to adjust his/her face in the right distance.

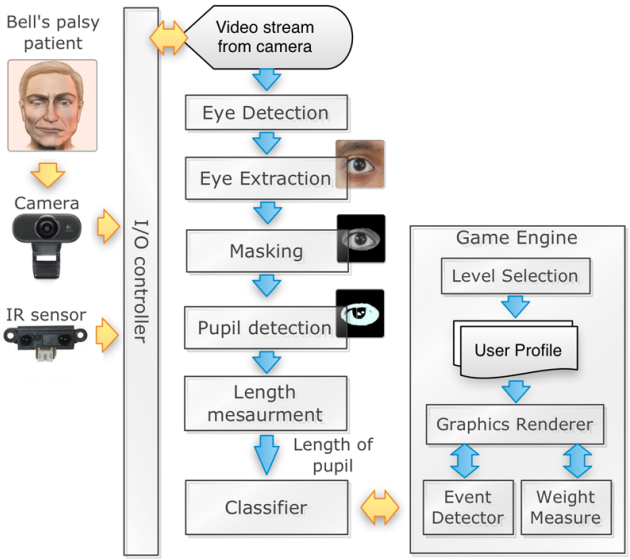


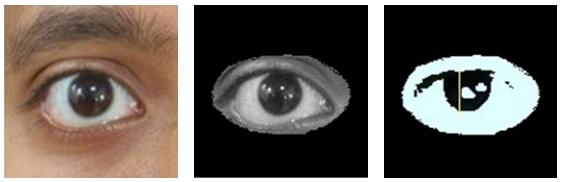
Fig. 2: Proposed gaming architecture for Bell's palsy patient.

A. Eye Extraction Module

Video is captured from the camera and extracted the video frames for eye detection. The eye is detected from the video frame by OpenCV [13] frontal haarcascade classifier. Haarcascade classifiers are XML files that contain information for object detection like face detection, eye detection etc. After the detection the eye is extracted from the face image. Later on the extracted eye image is scaled and cropped to 125*125 pixels image for masking.

B. Eye Image Masking Module

As different people have different skin and eye color so in order to get better result the system mask the cropped eye image. In the masking process the system will remove unnecessary skin parts, eyebrows and the noise. The skin parts will remove by comparing each pixel with human skin color. As human color is different so in our system has the facility to recognize the current user face color during face recognition stage. The system default face color is $RGB(239, 208, 207)$. In the final stage we perform some image processing operations to separate the pupil from the white part of the eye and from the skin. We're converting all the pixel colors that are greater than RGB (55, 55, 55) to the color RGB (224, 255, 255). Rest of the pixel colors is closer to the pupil colors.



(a) Cropped eye (125*125 pixel) (b) Gray scale and masking (c) Length measurement

Fig. 3: Different stages of eye image processing.

C. Pupil Length Measurement Module

After masking the image the system will measure the length of pupil. As in the masking stage most of the skin and noise are removed so the maximum connected colors that are not RGB (224,255,255) indicates the pupil length and this length value can be used to identify how much open the eye in-front the camera.

TABLE I: Properties of the videos that are used in the experiment.

	Interval (In pixel)	Openness (In percentage)
1	0-10	0%
2	11-19	25%
3	20-29	50%
4	30-39	75%
5	40 and above 40	100%

D. Classifier Module

Once the pupil length is measured the system determine the openness of the eye comparing with some predetermined values that are listed in table I. And based on those values, system will categories eye openness into five classes 0, 25, 50, 75 and 100 percent. As human eye is very responsive and the distance between fully close and fully open is very short so we discarded the intermediate classes. Following algorithm 1 is used to calculate the openness value.

Algorithm 1: Eye openness value calculation

```
Output: Eye openness array : g[]
begin
    Initialize window with OpenGL;
    while inside the glutmainloop do
        Draw initial bar for the game using workOnRectangle();
        while face exist in current image frame do
            frame ← Capture video frame using cvCaptureFromCAM();
            eyes[] ← IdentifyEye (frame);
            foreach eye e ∈ eyes do
                m ← maskImage(e);
                m_w ← Identify and transform color of the white portion
                of the eye;
                m_l ← Perform the length measurement of the pupil usign
                m_w;
                c ← classifyOpenness(m_l);
                g[] ← Pass the openness value c into method based on
                openness stages draw rectangle and ball position;
    return g[];
```

E. Game Engine Module

The game engine receives eye openness value from the classifier and makes use of them to control the game elements. The game has two different objects ball and a horizontal bar. Every certain amount of time a ball will fall down from the top to the bottom bar. If patient can open his/her eyes 100 percent the bar will be balanced horizontally and the ball will be in the middle. But if eyes are not open 100 percent then the ball will start to fall on the left or right according to the less openness. And if eyes are closed that means eye's openness is 0 percent the ball will fall off the bar and bar will be horizontally balanced. The game point will calculate based on the number of ball remain in the bar in a certain amount of time. So the patient concentration will be on the game point rather than exercise the eye. He will always try to keep maximum number of ball on the board i.e. keep the board horizontally balance.

IV. Implementation

For the purpose of checking the validity of our proposed method we have developed a prototype based on the proposed system architecture. We have developed both our module which combined to reflect the whole functionality of the exercise effect. This prototype is based on desktop environment. The environmental setup is illustrated in Figure 4.

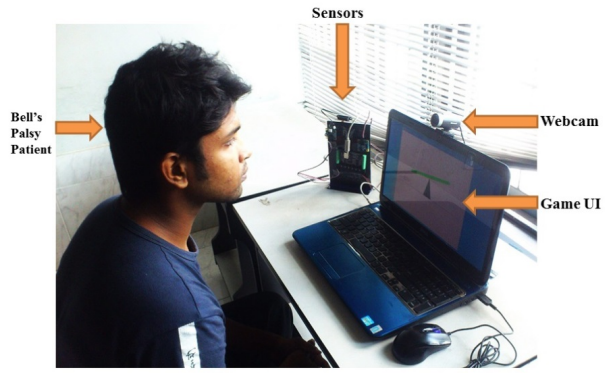
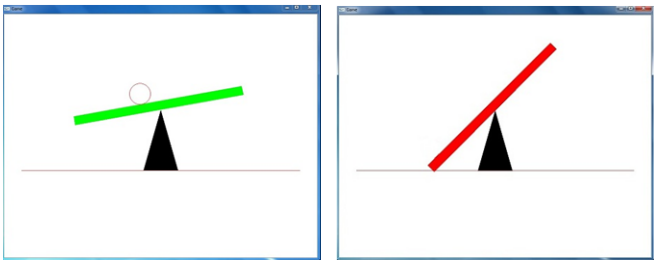


Fig. 4: The environmental setup where a patient is sitting in front of our prototype for exercise.

We developed our prototype system using OpenCV that enable us to interact with the camera. This portion is designed to use the primary camera, also to use the secondary camera if attached. The priority is set to secondary camera as high definition camera can be added on the system. We used a high definition Microsoft LifeCam HD3000 series to produce clearer picture so that noise is reduced in a high factor. Haarcascade that we choose to detect eye is the 'haarcascade_eye.xml', a XML file containing data about eye detection which comes with OpenCV. We used Phidget 1101_0 - IR Distance Adapter & Phidget 3520_0 - Sharp Distance Sensor to measure the distance of the subject. Subjects distance is ensured in between 12-16cm with this sensor. We also measured the environment light to ensure proper light availability with Phidget 1127 - Light Sensor. Light was kept in between 300-400lx for proper video capturing and avoiding shadows on eye portion. We used a Phidget interface kit to connect between sensors and computer. Open Graphics Library (OpenGL) is used to represent the feedback UI which is an interactive game. Besides this we used OpenGL physics engine for collision detection among balls and the board. The game UI is depicted in Figure 5. We used Visual Studio 2010 with .NET Framework 4 to develop our prototype.



(a) Bar position in 75% (b) Bar position in 0%

Fig. 5: Game UI and the bar position.

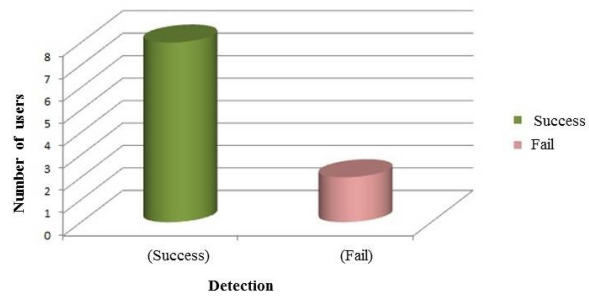


Fig. 6: Pupil's length measurement performance

V. Evaluation and Result

The goal of our evaluations is to determine the performance of our technique. For this reason, we have invited 18 healthy male subjects (including 10 real Bell's palsy patients), age 25-38, to participate in our experiments. These patients were found through local hospitals and known people. Every subject completed each of the exercises three times over 3 days. In the experiment each person invited to sit in front of the computer and keeping their face 12-16 cm away from the camera. Then we take their eye measurements through different shape. We asked them to open their eye naturally or using their hand and then gradually closing them. We took 5 stage images by which we defined pixel intervals for which openness percentage is defined. The environment was proper lighting condition that required for smoothening the video capturing for OpenCV and also with plain background to remove initial noise from the image. This plain background helps to avoid false positive eye detection. High definition cameras will give more accurate result in eye detection than general web-cams.

We have performed 2 types of measurement to evaluate the performance and the suitability of the proposed approach. In the first phase we recorded and demonstrated the success and failure for the pupil's length measurement which is a vital feature of our proposed architecture. And in the second phase we have shown the qualitative analysis by studying different aspects of prototype's feature. These are described in the following section:

A. Pupil's length measurement performance

In figure 6, we have shown the pupil's length measurement performance based on testing the prototype to 10 users. We have chosen 10 users at random and tested the eye detection and length measurement feature with them. We have seen if the system can measure the pupil correctly or the line that we draw to measure the length gone in wrong direction which is interpreted as failure. 2 different colors are used to visualize the success and failure. On the left the numbers represent the number of users.

Based on this performance we calculated the percentage of success, η by equation.

$$\eta = \frac{Number\ of\ successful\ measure}{Total\ number\ of\ users} = \frac{7}{10} = 0.7 \quad (1)$$

According to the graph our system successfully measure 70% user's pupil successfully.

B. Qualitative Measurements

To evaluate the user's quality of experience with the prototype and to justify the suitability of our proposed approach we have performed few qualitative measurement studies. The qualitative analysis is performed by studying different usability aspects of the proposed system. We have incorporated the usability evaluation guidelines [14] to qualitatively measure our proposed system and designed our tests accordingly. Before performing the usability test we designed a test plan where we defined our evaluation objectives, developed questions for the participants, identified the measurement criterion and decided upon the target users of the system.

The test took place at our university laboratory with ten (10) participants comprising of different age groups. Four (4) of the participants are in age group 25-30, five (5) of them are in age group 31-35 and the rest one (1) is in age group 36+. At a time, the selected volunteers were told to play the game and perform the exercise. After each session we have asked them to answer the following questions. The answers to the questions are in the range of 1-5 (the higher the rating, the greater the satisfaction) on a Likert scale [15].

TABLE II: Usability test questions.

#	Question
Q1	Is the system easy to understand than procedural system?
Q2	Is the system easy to learn and adapt?
Q3	The user is more at ease using this system?
Q4	The user do not need to actively use other body parts like traditional physiotherapy.
Q5	It is easy to measure the users progress using this method than the active physiotherapy system.

Figure 7 shows the user's responses for each given assertions. It is worth mentioning that more than 70% of the users would like to use the enhanced system. Overall the 75% users were also satisfied with the gaming approach based physical exercise. 70% patient expressed their willingness to exercise with this passive mode which increased their urge to exercise more and more than conventional method.

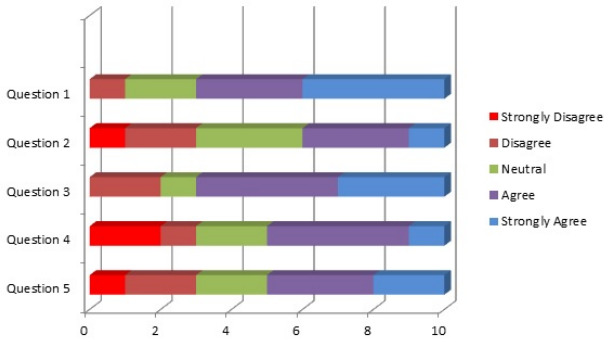


Fig. 7: User response for the usability test questions that are listed in Table II.

VI. Conclusion

In this paper we presented a system for the Bell's palsy patients to help them in passive exercise. The exercise method will be provided in the form of computer game so the user will be able to perform exercise without the hardship that comes with active exercise. We presented the implementation details of a preliminary prototype and performed usability studies on real patients. While we tried to address as many issues as possible related to the design of a prototype for passive exercise for Bell's palsy patient, there are still some topics that we did not treat or that could be treated in more detail. There are some limitations in our prototype tool. One of the biggest limitations of our system was currently the eye detection which we have done using OpenCV and which is not so much intelligent that it can recognize all types of eyes correctly and sometimes miscalculates other things as eyes. Another limitation is the synchronization time between the game board movements with the quick eye movement. If user open or close the eye very quickly then the system fails to recognize the correct openness of the eye. Besides this limitation, we believe that our proposed techniques for passive exercise for Bell's palsy patient will remain as our motivation for further research in this area.

References

[1] D. H. Gilden, "Bells palsy," in *New England Journal of Medicine*, vol. 351, 2004, pp. 1323–1331.

[2] A. Alamri, J. Cha, and A. El Saddik, "Ar-rehab: An augmented reality framework for poststroke-patient rehabilitation," in *IEEE Transactions on Instrumentation and Measurement*, vol. 59, October 2010.

[3] F. Anderson, M. Annett, and W. F. Bischof, "Lean on wii: Physical rehabilitation with virtual reality and wii peripherals," in *Stud Health Technol Inform*, vol. 154, 2010, pp. 229–234.

[4] C. Balance and A. Duel, "The operative treatment of facial palsy," *Arch Otolaryngol*, vol. 15, p. 170., 1932.

[5] B. palsy, "http://en.wikipedia.org/wiki/bell's_palsy."

[6] B. P. F. Sheet, "National institute of neurological disorders and stroke," in *Archived from the original on 8 April 2011*, 2003-04.

[7] K. Adour, F. Byl, R. J. Hilsinger, Z. Kahn, and M. Sheldon., "The true nature of bells palsy: analysis of 1,000 consecutive patients," in *Laryngoscope*, vol. 88, 1978, pp. 787–801.

[8] M. Sajadi and S. Tabatabaie, "The history of facial palsy and spasm: Hippocrates to razi," *Neurology*, vol. 77, p. 174182., July 2011.

[9] P. Z. OLSEN, "Prediction in recovery in bells palsy," in *Acta Neurologica Scandinavica*, vol. 52. Blackwell Publishing Ltd, 1975, pp. 1–119.

[10] D. Pitts, K. K. Adour, and R. L. Hilsinger, "Recurrent bell's palsy: Analysis of 140 patients," in *The Laryngoscope, John Wiley & Sons, Inc*, vol. 98, pp. 535–540.

[11] T. Shafshak, "The treatment of facial palsy from the point of view of physical and rehabilitation medicine," *Europa Medicophysica*, vol. 42, pp. 41–48, 2006.

[12] M. J. A. Jannink, G. J. V. D. Wilden, D. W. Navis, G. Visser, J. Gussinklo, and M. Ijzerman, "A low-cost video game applied for training of upper extremity function in children with cerebral palsy: a pilot study," in *CyberPsychology & Behavior*, vol. 11, no. 1, 2008, pp. 27–32.

[13] OpenCV, "http://en.wikipedia.org/wiki/opencv," Tech. Rep.

[14] D. Chisnell, "Usability testing: Taking the experience into account." in *IEEE Instrumentation Measurement Magazine*, vol. 13(2), 2010, pp. 13–15.

[15] wikipedia, "Likert scale, , http://en.wikipedia.org/wiki/likert," Tech. Rep.

DBF-MLT: NP-Completeness of Specialized Voice Service in Networked Virtual Environment a Peer-to-Peer Approach

SK Alamgir Hossain
Computer Science and Engineering Discipline
Khulna University
Khulna, Bangladesh
Email: alamgir@cseku.ac.bd

Tamanna Tabassum
CFS Limited
Khulna, Bangladesh
Email: tamanna@thecodeandfix.com

***Abstract*—Several researchers have successfully developed realistic technique for specialized voice service in networked virtual enthronement (*NVE*) like Second Life. Specialized voice service in virtual environment enables users to speak to each other naturally and seamlessly. The spatial 3D avatar location and the voice direction in the virtual environment are used to generate realistic sound. Although with this specialized audio or voice streaming in virtual environment the real user can more easily identify who is talking, if several people are taking simultaneously. But this is a challenging task to stream the voice or audio data to the users' machine by considering the user bandwidth limit and the tight latency constraint and this can be formalized as a NP-complete problem. In this paper, we presented the proof of NP-completeness of the specialized audio streaming problem called *DBF-MLT*.**

***Keywords*—NP-Completeness, Networked Virtual Environment, Second Life, Specialized Voice Service**

I. Introduction

Plato introduced the concept of virtual reality in 380 BC, through the famous writing *the cave* [1]. Technology realizes Plato's writing of the *the cave* through networked virtual environment. Networked virtual environment, is a computer simulated environment that reproduces a real world where human control their activities inside the world through their digital representation called *avatar*. One of the most popular and rapidly spreading examples of networked virtual environment is Linden Lab's *Second Life*[1]. In Second Life, once connected the users can view their avatars in the 3D virtual world and they can participate in real-time in task-based games, play animation, communicate with other avatars ([2] [3]) through instant messaging and specialized voice services. The goal of this specialized voice services provided in a virtual world is to enable users to speak to each other naturally and seamlessly. The most important disadvantage of existing audio service in networked virtual environment is that they are not designed the spatial information of a virtual avatar. With specialized voice service the real user can more easily identify who is talking, if several people are taking simultaneously.

[1]Second Life, http://secondlife.com

In our real life we are familiar with this type of activity. If several people are talking from several distance then the sound receive by a person depends on different factors like sound level, distance, frequency, direction etc. Similarly, for specialized voice service the audio stream depend on the avatar virtual position, direction and the user bandwidth limit.

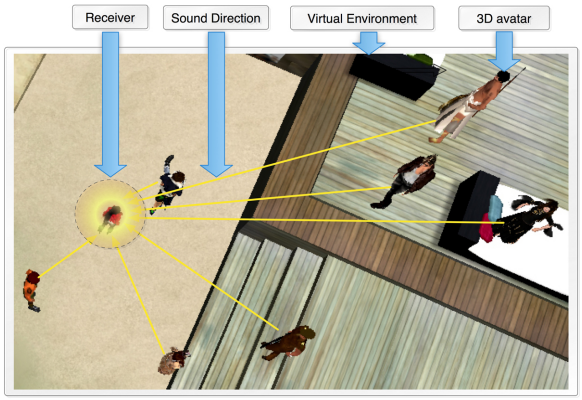


Fig. 1: Relative Distance of Avatars in the Networked Virtual Environment and the physical positions of real users. The specialized voice service depends on the avatar virtual position, direction and the user bandwidth limit.

By utilizing a peer-to-peer based approach which can alleviate the scalability problems caused by the client-server design the audio streams disseminate using application level multicast trees built on overlay networks that take into consideration heterogeneous bandwidth limits of users and tight latency constraints. Because the upload and download bandwidths of each node in a peer-to-peer network are limited and heterogeneous, the task to effectively utilize the network resources when disseminating several separate audio streams while keeping the overlay latency bounded becomes very challenging. Some unsuccessful receivers may not be able to receive the audio stream that they supposed to receive. Similarly some successful receivers may have higher delay

or latency because of the non-optimal dissemination path of the audio stream. Hence, the problem aims to:

- Maximize the number of successful receivers of corresponding audio streams and
- Minimize the latency for those successful receivers.

Specialized voice service and audio rendering technique in networked virtual environment is not new, many researcher around the world are working on different services specially realistic audio streaming service in networked virtual environment. A notable work is [4], the author presented the audio rendering technique in virtual environment when the audio sources are moving in different directions. They used the spatial level of details for the 3D audio rendering in complex virtual environment. Based on their studies they argued that by incorporating simple spatial information the audio rendering can be noticeable improve. Another, spatial audio rendering technique is [5], where the authors nicely introduced the audio streaming problem in the networked virtual environment and the optimized technique for rendering a sufficient number of dynamically moving sound sources in the virtual environment. Yamazak et. al [6] also explore the use of spacial audio streaming and the user's experience in multiuser virtual environments. They developed a system called *chatspace* for exploring spatial audio conferencing functionality in multiuser virtual environments.

Although various audio streaming technique have been devised but considering the bandwidth and processing speed peer-to-peer based data delivery is better that the client server based approach [7]. The author presented that the very nature of 3D streaming necessarily places a heavy burden on the streaming server providing the service and peer-to-peer based approach is more useful for distributed networked virtual environment.

Audio streaming performance not only depends on the spatial direction of voice but also depends on the server load. Lui et. al [8] presented how to assign workload to different servers in virtual environment and at the same time, reducing the overall workload and the communication cost. They showed this partitioning problem as a NP-complete problem and discussed several heuristics to improve the service.

Our contribution in this paper is two-fold. First, we provide the problem formulation of an interesting problem in network virtual environment. Second, we presented the NP-completeness proof by reduction of an existing NP-complete problem.

The remainder of this paper is organized as the following. At first, in Section II we formulate the problem statement. Next in section III we presents the NP-completeness proof of *DBF-MLT* problem. At the end we provide conclusion of the paper in Section IV and state some possible future work directions.

II. Problem Formulation

The users (avatars) in a virtual environment can move in the system at will and any two users can communicate with each other when their distance is less than *r*, the radius of a node's area of interest (*AoI*). Every user can speak anytime and at any location. Each user *j* has a limited upload bandwidth u_j^{max}. The upload capacity is measured in units, where one unit is defined as the required bandwidth for a single audio stream and measured in bits/s. We model the system as a digraph *G = (V, E)*. Where, *V* is the set of nodes representing users (avatars) in the system and *E* is the set of edges. An edge *e(i, j)* indicates that node n_i is willing to send (if n_i is a speaker) or forward an audio stream to n_j.

When a node starts speaking, all its neighbour nodes are candidate receivers for the audio stream sent from the speaker. Due to the bandwidth limits of the neighbour nodes and the speaker, only a fraction of the candidate receivers, called successful receivers, may actually be able to receive the stream. All other nodes outside of this range will not be able to hear or forward the audio stream. When a node n_i starts to speak, it will construct a temporal multicast tree T_0^i rooted at n_i. All candidate receivers are members of T_0^i, that is, T_0^i is initially constructed without considering the bandwidth limits of nodes. T_0^i is a Shortest Path Tree (*SPT*) which is a combination of shortest paths from the speaker to all nodes within its *AoI* (see figure 2 for more details).

From figure 2, it is clear that we need to construct a forest that consists of multicast trees rooted at all the speakers (avatar), provided that we have to ensure the bandwidth limits of nodes and latency or delay. We define the problem as constructing a forest of degree-bounded minimum latency trees called here, *DBF-MLT*.

III. NP-Completeness of DBF-MLT

We can formulate our *DBF-MLT* problem as a decision problem. Here we have a set of nodes (avatars) n_1, n_2, ... ,n_m and a distance matrix (d_{ij}) giving the distance between avatar n_i and n_j. Consider the following notation in Table I.

TABLE I: Symbols and notations

Symbol	Meaning
L(j, k)	the latency of the path from node n_j to node n_k
S	the set of speakers
R_i	set of successful receivers of speaker n_i

$$Tour,\ \zeta = \frac{1}{R^S} \sum_{j \in S} \sum_{k \in R_j} L(j,k) \tag{1}$$

$$|R^S| = \sum_{i \in S} |R_i| \tag{2}$$

We wish to construct a tour that minimizes Eq 1. In this equation if the value $|R^S|$ is maximum then ζ will be minimum. We also need to satisfy the upload bandwidth limit of user *x*, i.e. $u_x \leq u_x^m, m \in V$. So our question is we have multi-cast trees rooted at all the speakers and is the tree is optimal with at most the total latency is *B*, i.e. is it satisfy our needs (maximize the number of receiver

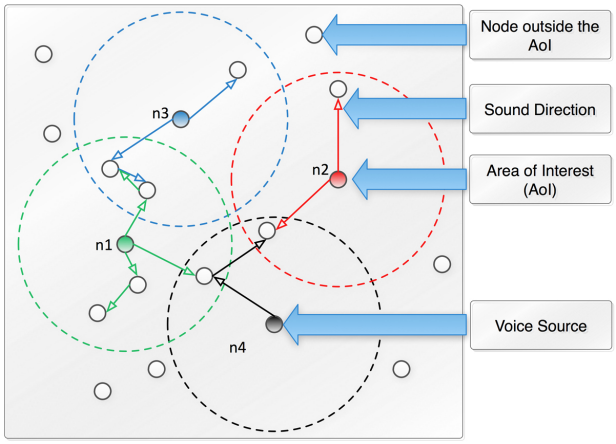


Fig. 2: Audio Streaming Trees within the AoI of Four Speaker Avatars, n1, n2, n3 and n4.

avatar with minimizing the latency by considering the user bandwidth limit).

The process of proving NP-completeness ([9] [10]) of *DBF-MLT* (degree-bounded minimum latency trees with proper bandwidth limit) passes through the following steps :

- Show *DBF-MLT* is in *NP*.
- Select a known *NP-complete* problem such as Freeze-Tag Problem (*FTP*) [11] [12].
- Construct a reduction *f* from *FTP* to *DBF-MLT*.
- Prove the reduction *f* is in polynomial time solvable.
- Finally, prove that for all $I \in D_{DBF-MLT}$, $I \in Y_{FTP}$, iff $f(I) \in Y_{DBF-MLT}$.

Following are the above five steps for proving the NP-Completeness of audio streaming problem in *NVE* that discussed so forth. The first step is to show *DBF-MLT* problem is in *NP*, because it can be decided by a nondeterministic Turing machine [13], in time $O(n^2)$. This machine, on an input containing a representation of the *DBF-MLT* instance, goes on to write an arbitrary sequence of symbols, no longer than its input. When this writing stops, the machine goes back and checks to see whether the string written is the representation of a permutation of the nodes, and, if so, whether the permutation is a traverse with latency B or less. Both tasks can easily be carried out in $O(n^2)$ time. If the string indeed encodes a traverse cheaper than B, the machine accepts; otherwise, it rejects. This machine indeed decides *DBF-MLT*, because it accepts its input if and only if it encodes a "yes" instance of *DBF-MLT*. If the input is a "yes" instance, this means that there is a tour of the nodes with latency B or less. Therefore, there will be a computation of this machine that "guesses" precisely this permutation, checks that its cost is indeed below B, and thus ends up accepting. It does not matter that other computations of the machine may have guessed very costly tours, or plain nonsense strings, and will therefore reject; a single accepting computation is all that is needed for the machine to accept. Conversely, if there is no tours of latency less than B, there is no danger that a computation will accept: All computations will discover that their guess is wrong, and will reject. Thus, the input will be rejected. This will fulfill the first step that *DBF-MLT* problem is in *NP*.

The next step is to choose a known *NP-complete* problem and reduce the problem to our *DBF-MLT* problem. For the reduction here Freeze-Tag Problem (*FTP*) [11] [12] is taken as a known *NP-complete* problem. Freeze-Tag is a very popular children's tag game. A person who contain "it" can tag the other and the tagged player become "frozen". Other player who are not already "frozen" can unfreeze a "frozen" player. The game end when all the players are become "frozen". This tag game can be modeled in case of robot awake problem. Given a complete graph *G' = (V', E')*, the Freeze-Tag Problem can be modeled as a set of robots *V'*, in which initially one robot *S'* awake and others are *R'* asleep. The robots which are awake can be select *B'* sleeping robots and awaken them, provided that the task must be completed in optimal awakening schedule. Once awaken, each new robot is available to assist in rousing another set of *B'* robots. There exists latency between each pair of robots. Thus, a solution to the *FTP* can be described by a wake-up tree *T* which is a directed *B'-ary* tree (means is a rooted tree in which each node has no more than *B* children) rooted at *S'*, spanning all robots *R'*. The objective is to minimize the make span of *T*, i.e., the time *t'* when the last robot awakens.

Now, let the formulation of digraph *G = (V, E)* as shown in Figure 2 in *DBF-MLT* be identical to *G'* in *FTP*, and let the audio streaming rate be one (1) bits/s and all the nodes *V* in *G*, the maximum upload capacity is $B' = u_j^{max} bits/s$, means every node can stream audio up to *B'* childrens. So the resulting topology becomes a *B'-ary* spanning tree *T*, i.e. the resulting tree is a rooted tree in which each node has maximum *B'* children. The maximum end-to-end audio streaming delay (*t*) is the latency of the last peer in *T*. So, the optimal solution (maximize the number of receiver avatar with minimizing the latency by considering the user bandwidth limit) to the *DBF-MLT* problem can also solve the *FTP* problem optimally (i.e. minimize the make span of *T*). In other words, the *FTP* will have a tree *T* in *G'* with minimum delay of min *t'* if and only if the same *T* in *G* provides the minimum-delay spanning tree and the resulting min-max latency of *DBF-MLT* equals min *t'*. So the reduction is complete and clearly the simple reduction can be done in polynomial time.

Now suppose that unsuccessful receiver set U_i (all nodes that have no more available in degrees) is removed from the list of nodes (users). The multicast tree is constructed based on the remaining nodes, and rooted at the sender. In order to achieve low latency we construct a minimum latency tree (*MLT*), constructed by combining all shortest paths from the sender to its neighbours and for maximize the number of successful receivers a heuristic called *MulticastTreeAdjust* used to fulfill our requirements. From this multicast tree which is also a minimum latency tree the audio stream transfer from the sender to the receiver in a minimum delay d. If the user j

bandwidth limit is $u_j^{max} bits/s$ and if the audio stream rate is one $bits/s$ then clearly the sender has at most u number of children or it is a *u-ary* tree. If we consider the tree node as a robot then surely the *u-ary* tree is an instance of a *FTP*, where the robot which is awake will select u sleeping robots and awaken them. Once awaken, each new robot is available to assist in rousing another set of u robots. As the tree is a minimum latency tree so wake-up tree *T'* of *FTP* rooted at the sender must be minimize.

So we have proved both sides of equivalence. A *DBF-MLT* of latency less than *B* exists in *G* if and only if a Freeze-Tag Problem (*FTP*) of *B'-ary* tree exists in *G'*. Since we know that Freeze-Tag Problem (*FTP*) is an *NP-complete* problem, *DBF-MLT* is also *NP-complete*. This part proves the last step that for all instance *I*, $I \in Y_{FTP}$, iff $f(I) \in Y_{DBF-MLT}$. Thus all steps of showing *NP-Completeness* of *DBF-MLT* has done. So *DBF-MLT* is in *NP-Complete*.

IV. Conclusion

In this paper, we presented the proof of NP-completeness of specialized audio streaming in networked virtual environment. With specialized audio or voice streaming in virtual environment the real user can more easily identify who is talking, if several people are taking simultaneously. Like the real life the streaming service depends on the virtual avatar position, the direction of sound and the user bandwidth limits. For the proof we used the well-known NP-complete problem called Freeze-Tag problem. Although, it is hard to measure some of the parameters in the virtual world for example the avatar voice direction. In this work we investigate the preliminary characteristic of the audio service; we would like to address the issues more details in future. However, we believe that our proof will remain as a motivation for further research in this area.

References

[1] R. L. Huard, *Plato's Political Philosophy: The Cave*. Algora Publishing, 2007.

[2] S. Hossain, A. Rahman, and A. El Saddik, "Interpersonal haptic communication in second life," in *IEEE International Symposium on Haptic Audio-Visual Environments and Games (HAVE)*, Oct 2010, pp. 1–4.

[3] S. A. Hossain, A. Rahman, and A. El Saddik, "Measurements of multimodal approach to haptic interaction in second life interpersonal communication system," *IEEE Transactions on Instrumentation & Measurement*, vol. 60, no. 11, pp. 3547–3558, Nov 2011.

[4] N. Tsingos, E. Gallo, and G. Drettakis, "Perceptual audio rendering of complex virtual environments," in *ACM SIGGRAPH 2004 Papers*, ser. SIGGRAPH '04. New York, NY, USA: ACM, 2004, pp. 249–258. [Online]. Available: http://doi.acm.org/10.1145/1186562.1015710

[5] M. Naef, O. Staadt, and M. Gross, "Spatialized audio rendering for immersive virtual environments," in *Proceedings of the ACM symposium on Virtual reality software and technology*, ser. VRST '02. New York, NY, USA: ACM, 2002, pp. 65–72. [Online]. Available: http://doi.acm.org/10.1145/585740.585752

[6] Y. Yamazaki and J. Herder, "Exploring spatial audio conferencing functionality in multiuser virtual environments (poster session)," in *Proceedings of the third international conference on Collaborative virtual environments*, ser. CVE '00. New York, NY, USA: ACM, 2000, pp. 207–208. [Online]. Available: http://doi.acm.org/10.1145/351006.351051

[7] S.-Y. Hu, "A case for 3d streaming on peer-to-peer networks," in *Proceedings of the eleventh international conference on 3D web technology*, ser. Web3D '06. New York, NY, USA: ACM, 2006, pp. 57–63. [Online]. Available: http://doi.acm.org/10.1145/1122591.1122599

[8] J. Lui, M. Chan, O. So, and T. Tam, "Balancing workload and communication cost for a distributed virtual environment," in *Advances in Multimedia Information Systems*, ser. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 1998, vol. 1508, pp. 130–135. [Online]. Available: http://dx.doi.org/10.1007/3-540-49651-3_13

[9] M. R. Garey and D. S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness*. New York, NY, USA: W. H. Freeman & Co., 1979.

[10] M. Sipser, *Introduction to the Theory of Computation*, 1st ed. International Thomson Publishing, 1996.

[11] E. M. Arkin, M. A. Bender, S. P. Fekete, J. S. B. Mitchell, and M. Skutella, "The freeze-tag problem: how to wake up a swarm of robots," in *Proceedings of the thirteenth annual ACM-SIAM symposium on Discrete algorithms*, ser. SODA '02. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 2002, pp. 568–577. [Online]. Available: http://portal.acm.org/citation.cfm?id=545381.545457

[12] M. Sztainberg, E. Arkin, M. Bender, and J. Mitchell, "Theoretical and experimental analysis of heuristics for the "freeze-tag" robot awakening problem," *Robotics, IEEE Transactions on*, vol. 20, no. 4, pp. 691 – 701, 2004.

[13] Wikipedia, "Non deterministic turing machine, http://en.wikipedia.org/wiki/non-deterministic_turing_machine," Tech. Rep., Last Accessed, March 2011.

Neural net based complete character recognition scheme for Bangla printed text books

SK Alamgir Hossain
Computer Science and Engineering Discipline,
Khulna University, Khulna, Bangladesh
alamgir@cseku.ac.bd

Tamanna Tabassum
CFS Limited
Khulna, Bangladesh
tamanna@thecodeandfix.com

Abstract
In this paper we propose a neural net based characters recognition scheme for Bangla printed text books. There are a lot of scientific literature, novels, magazines and books etc that are written in Bangla language. More than 400 million people use Bangla language. Most of the library and educational institutions want to keep copy of the books in a digital format. For storing those books in digital text format we need a good character recognition schemes by which we can convert the scanned text book images to editable texts. The key contribution of our research highlights this issue. We propose four main stages namely preprocessing, segmentation, training-recognition and post-processing. In the beginning the input book images preprocessed by rotation, scaling, binarization and noise elimination. The binarized image is then segmented and extracted into individual characters that are trained and recognized by an artificial neural network. Finally, the process ends by reconstructing the text in the post processing stage.

Keywords - Binary Image, Boundary Fill, Character Recognition, Neural Network, OCR

I. INTRODUCTION

The significance of optical character recognition (OCR) [7] increases day by day. Manual conversion of documents is too slow and expensive. Many commercial companies have developed OCR systems for English, Japanese, Chinese and other European languages. OCR in Bangla language is comparatively new and very few research works are available in this era. OCR in Bangla language is more complex than other languages as Bangla has a large number of single characters as well as compound characters. Table I shows the Bangla alphabet characters. Two or more of this alphabet characters may combine to form complex shaped characters called conjunctive characters. Each character requires different level of detection mechanism. Detection of simple characters such as ক খ গ ঘ ঙ চ ছ is easier than the detection of characters such as ট ঠ ই ি ী ্ব. In this paper we present a method of OCR system which can work efficiently specially for those books which are written in printed text. The main contribution of our method is its rotation invariant capability. Our method also performs well for recognizing conjunctive characters.

TABLE I: BANGLA ALPHABET CHARACTERS

Alphabet consonants	ক খ গ ঘ ঙ চ ছ জ ঝ ঞ ট ঠ ড ঢ ণ ত থ দ ধ ন প ফ ব ভ ম য র ল শ ষ স হ ড় ঢ় ক্ষ য় ৎ ং ঃ
Alphabet vowels	অ আ ই ঈ উ ঊ ঋ এ ঐ ও ঔ
The vowel marks	া ি ী ু ূ ৃ ে ৈ ো ৌ ঁ

In our proposed method, first the scanned grayscale image is converted into black & white image, afterwards we perform rotation and noise removal operations. These preprocessing steps are described in Section IV. The segmentation scheme is presented in Section V. The segmented characters are trained and recognized by artificial neural network (ANN); this process is described in Section VI. Further, in Section VII we describe the methodology by which the recognized characters are combining to rebuild the original text. We performed some experimental study to show the performance of our scheme in section VIII.

II. RELATED WORKS

A.R. Forkan et al. [1] describe Bangla OCR system only of conjunctive characters. In their work they suggested a flexible matching between sample data and training data components applying Multi-layered Feed-Forward ANN. The accuracy level of some special fonts such as RinkiyMJ (22pt) shows result of 87.56% recognition. They do not consider complex regular characters such as ট ঠ ই ি ী ্ব. A. Dutta et al. [2] describe both isolated handwritten and printed OCR system for Bangla. A detail description of the characteristics of Bangla text is discussed. They use a curvature related feature for characterizing the strokes which constitute the characters. One main limitation of the scheme is with their method; it is not possible to recognize composite characters.

S.M. Shoeb Shatil et al. [3] describes segmentation-free optical character recognition system for Bangla using a Kohonen network. The methodology presented in this paper assumes that the OCR preprocessor has minimally segmented the input words into easily segmentable chunks, and presenting each of these as an image to the classification engine. The main problem of this work is that the recognition process is not rotation invariant. Md. Abul Hasnat et al. [4] presented a complete Bangla OCR for domain specific document images. The performance level for some domain is not satisfactory. M.A Sattar et al. [6] only discussed the segmentation stage. They do not mention any information about how to apply the segmented characters in an actual recognition process.

III. BANGLA OCR STEPS

In our proposed characters recognition approach (Fig.1.) we divided the whole system into a number of steps as the following:

a. Printed book images are taken directly from the scanner in grayscale mode. The gray scale image is then converted into binarized image.
b. The scanned image may rotate so the next step is to rotate the binarized image into its correct position.
c. If the image is not in the proper size, it will be resized through scaling. The next step is to make the image noise free in the noise elimination stage.
d. A next important step is the segmentation step which is divided into two parts: we segment the lines into words and the words into characters.
e. After segmenting the image into characters, we pass those gained to the training and recognition stage. In the training stage, we train the neural network using randomized data. After the training process is complete, actual character recognition process is carried out. The neural net will update its internal weights in case errors are found in the gained output.
f. In the final steps the recognized characters are reconstructed to the actual text format.

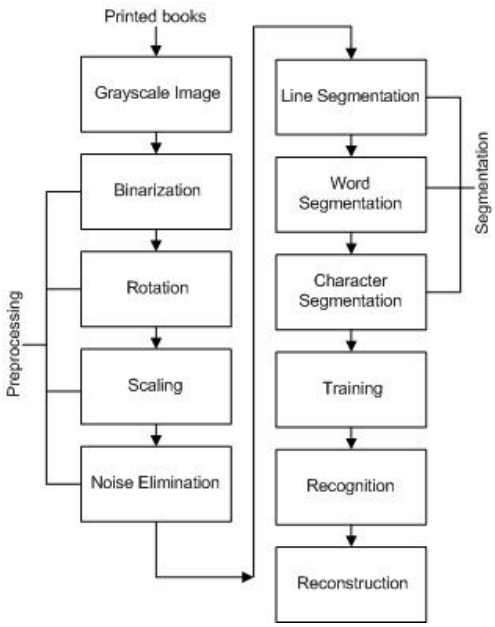


Fig. 1. Proposed Bangla OCR steps

IV. PRE-PROCESSING

In the first step a Bangla printed book's text image is obtained from a flat bed scanner in grayscale mode with 300dpi resolution. The scanned images may have some noise or be scanned in different sizes. So we need to preprocess the scanned images before sending them to the segmentation stage. Preprocessing stage is divided into four functions namely binarization, rotation, scaling and noise elimination.

A. Binarization

The scanned gray scale image is converted to a black & white image. In the grayscale image the pixel values are in the range from 0 to 255. In the black & white image each pixel has only one value either 0 or 1. A black pixel has value 0 and a white pixel has value 1. In Fig.2 the histogram of a gray scale image and its corresponding black & white image is shown.

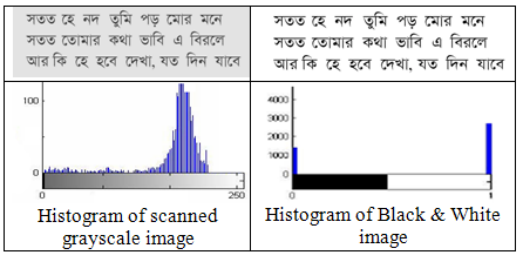


Fig. 2. Histogram of grayscale image (values from 0 to 255) converted Black/White image (two values: 0 and 1)

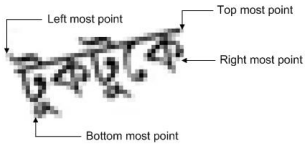


Fig. 3. Rotated image and its different points.

B. Rotation

Sometimes the scanned images may rotate due to the position of the book on the scanner. Most of the Bangla OCR fail to recognize the characters if the input image is rotated. In our proposed method before segmenting the images, we rotate it to its correct position. The rotation point of each word is depicted in Fig.3. The main challenge in this operation is to measure the exact degree of information by which the image needs to be rotated to its actual position. We have devised the following algorithm in order to perform proper rotation of the image.

Algorithm ROTATE_IMAGE (*img: IMAGE*)

*/*gets the input image and returns the number of degree that needs to be rotated so that the image comes to its correct position. leftPoint, rightPoint, topPoint, botPoint are local variables which represents the left, right, top and bottom points. LD, RD, dx, dy are local variables (See Fig. 3)*/*

```
1.  for every column (col) from leftPoint to rightPoint
2.      if image position (topPoint, col) is 0 then
3.          Set LD to (col – leftPoint)
4.          Set RD to (rightPoint – col)
5.      end if
6.  endfor
7.  for every row from topPoint to botPoint
8.      if LD greater than RD then
9.          Set dx to LD
10.         if image position (row, leftPoint) is 0 then
11.             Set dy to (row – topPoint)
12.             go to line 22
13.         end if
14.     else
15.         Set dx to RD
16.         if image position (row, rightPoint) is 0 then
17.             Set dy to (row – topPoint)
18.             go to line 22
19.         endif
20.     endif
21. endfor
22. output atan(dy/dx)*180/pi
```

C. Scaling

The rotated image may not be in the proper size, therefore before segmenting the characters we need to scale the image to an appropriate size. In our method we scale the images to

a size with an aspect ratio of 1 to 0.8. This is the ratio of standard letter size pages.

D. Noise Elimination

The scanning and or transformation steps of the image may add some noise. Before taking further steps we eliminate this noise. In our proposed method we have used averaging and linear smoothing filter for noise elimination.

V. SEGMENTATION

After preprocessing the image the text blocks are first segmented into lines, lines are then transformed into words and finally words transformed into individual characters.

A. Line Segmentation

Line segmentation is performed by scanning the input image horizontally from top to bottom. The first black pixel of a horizontal line denotes a starting boundary of a line. We continue to scan until we find a horizontal line which has no black pixel. This indicates the ending boundary of the line. Fig. 4 demonstrating the starting and ending boundary of different lines.

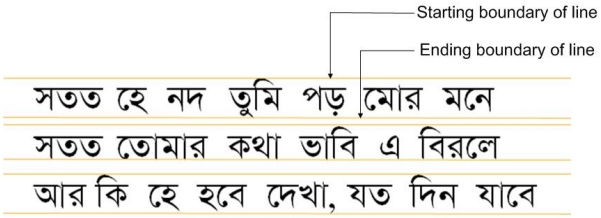


Fig. 4. Line segmentation.

B. Word Segmentation

After text line segmentation is performed, the system segments each line into words. For word segmentation each line is vertically scanned from left to right in order to detect a vertical line with at least one black pixel. This precondition denotes the point of a word. For detecting the ending point of a line we continue to scan vertically until finding a vertical line which has no black pixel. Fig. 5 depicts a text line and its different parts. In Bangla several characters are joined by a horizontal line called "*Mattra*" line. Mattra line is that row or set of rows where the number of black pixel is the maximum [11]. The upper zone denotes the portion above the mattra line, the middle zone covers the portion of basic and compound characters below the mattra line and the lower zone may contain where some vowel and consonant modifiers can reside. The imaginary line separating the middle and lower zone may be called the base line. Upper mattra line is the upper boundary of a line and lower base line is the lower boundary of a line.

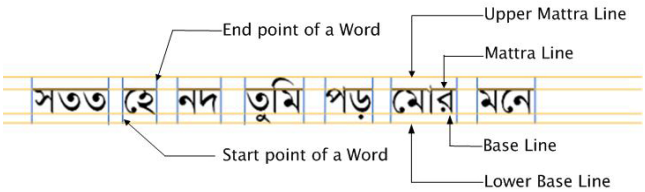


Fig. 5. Different parts of word segmentation.

C. Character Segmentation

The performance of an OCR mainly depends on the accuracy of the character segmentation stage. Character segmentation is performed in some sub stages: (a) removing the mattra line (b) segment those characters which is within the mattra line and base-line (c) segment the parts which are above the mattra line and under the base line. For removing the mattra line, we scan the image horizontally and count the number of black pixel in each row. All the horizontal pixel lines containing this maximum number of pixels or at least 90% of this maximum value is considered as mattra lines. For segmenting the character within the mattra line and the base line we scan vertically. The starting point of a character is the first column where the first black pixel is found. After finding the starting point of a character, we continue scanning until a column without any black pixels is found, which we consider as the ending point of the character being processed (see Fig. 6).

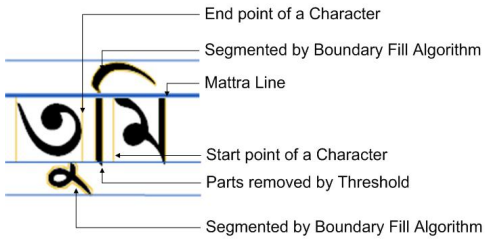


Fig. 6. Character segmentation.

To segment the upper portion and lower portion of the mattra line and base line we use the following BOUNDARY_FILL algorithm.

Algorithm BOUNDARY_FILL (*x, y, img:* IMAGE)
*/*the algorithm fill a region in image img from black to white color. Filling starts at point (x, y). It also copy the filled region into another image img2*/*

1. **if** pixel (*x, y*) in image *img* is black **then**
2. Copy the current pixel value to image *img2*
3. Reset the current pixel location with a white pixel
4. Call BOUNDARY_FILL with values (*x+1, y*)
5. Call BOUNDARY_FILL with values (*x-1, y*)
6. Call BOUNDARY_FILL with values (*x, y+1*)
7. Call BOUNDARY_FILL with values (*x, y-1*)
8. Call BOUNDARY_FILL with values (*x-1, y-1*)
9. Call BOUNDARY_FILL with values (*x-1, y+1*)
10. Call BOUNDARY_FILL with values (*x+1, y-1*)
11. Call BOUNDARY_FILL with values (*x+1, y+1*)
12. **end if**

VI. TRAINING AND RECOGNITION

For training and recognition, each character images is first converted to a 16x16 pixel image; the only feature extracted from the images is a 256-bit long vector, which is then trained or classified using an artificial neural network.

A. Training

Neural network has been trained by normalized feature vector obtained for each character in the training set. In our system the normalized vector is 256-bit long. Four layer neural networks have been used with two hidden layers for improving the classification capability of the neural network with minimum error tolerance rate. At first we will train our neural net with some randomized data. This data will update (hidden weights) based on the error found in the recognition stage.

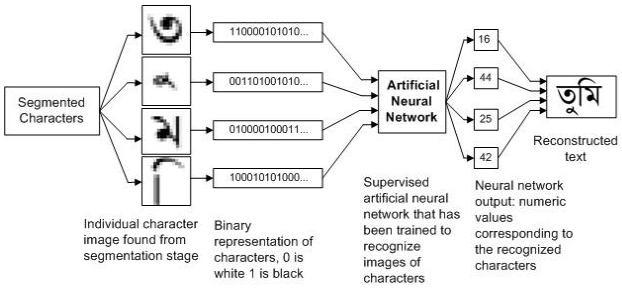


Fig. 7. Segmentation and Regeneration process.

B. Recognition

In the recognition phase of the network, a single iteration is enough to give the confidence value for each class of the character set. After applying the 256-bit long vector (Fig.7.) to the network, if the output of the network is very close to one of the characters with a certain acceptable tolerance then our system will output a numerical value representing the recognized number which then again joins to form the original texts. If the output is far apart from the threshold or the acceptable range from all the possible outputs, then our system cannot identify the character and return a numerical value 0, indicating an unrecognized character.

7. RECONSTRUCTION

After recognition of each character with the neural network we need to regenerate the original text. For reconstruction of the original texts we follow the Bangla word formation rules [9]. According to Bangla word formation rules if a vowel is immediate before the consonant then it will be a vowel marks like Fig. 8.

Fig. 8. Reconstruction the text.

VIII. EXPERIMENTAL RESULTS

We tested our OCR system to measure the performance, for this we scanned 32 pages from popular Bangla novel *Pather Panchali* [8]. There are about 920 lines and 10,400 words in our test case. We separately measure the line, word and character segmentation. We tested our code written in MATLAB. According to our test result our line segmentation success rate is about 98.8% for 920 lines. Our word segmentation success rate is 96.2 % for 10,400 words and the character segmentation success rate is 81% for 10,400 words. In our experiment we used the very popular true type Unicode Bangla font name *AdorshoLipi*. We also tested in some other popular Bangla true type fonts, the results are depicted in Fig. 8. From the result in Fig. 9. our method shows a good performance for most of the fonts except some fonts like *Suleka, Siyam Rupali*.

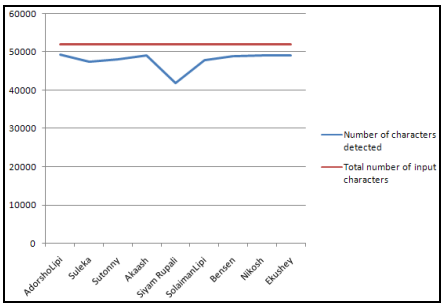


Fig. 9. Performance curve in different Bangla fonts.

We have compared our method with some other popular methods [3, 5], the results are shown in Fig.10.

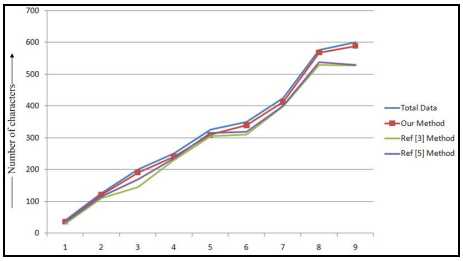


Fig. 10. Comparison graph with other methods.

We have also tested our method in different positive and negative angles and the percentage of success to calculate the exact angle is shown in Fig. 11.

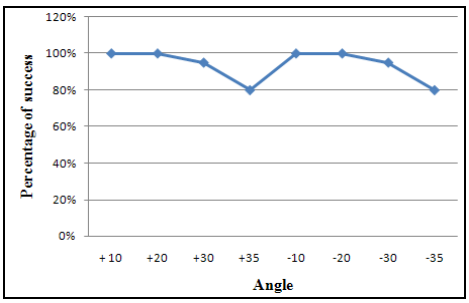


Fig. 11. Different angle and percentage of success.

IX. CONCLUSION

The main objective of this research is to develop a complete character recognition scheme for Bangla printed text books which will be rotation invariant with an acceptable performance rate. Based on our experimental results we can comment that our line and word segmentation method have satisfactory performance. Our character segmentation shows some error because there are some characters in Bangla for which recognition is very difficult. Moreover our method shows a good performance in different types of true type Bangla fonts. Based on the result study our research shows the potential to be used in Bangla text book digitization. We believe that the proposed method of character recognition scheme for Bangla printed text books will bring new direction in future research in this research area.

REFERENCES

[1] A.R. Forkan, S. Saha, M.M. Rahman, M.A. Sattar, "Recognition of Conjunctive Bangla Characters by Artificial Neural Network", *ICICT 2007*, 7-9 March 2007.

[2] A. Dutta, S. Chaudhury, "Bengali Alpha-Numeric Character Recognition Using Curvature Features", *Pattern Recognition*, Vol. 26, No. 12, pp. 1757-1770, 1993.

[3] S. M. Shoeb Shatil and Mumit Khan, "Minimally Segmenting High Performance Bangla OCR using Kohonen Network" , *ICCIT 2006, Proc. of 9th International Conference on Computer and Information Technology*, 2006.

[4] Md. Abul Hasnat, M.H Mumit Khan, "A High Performance Domain Specific OCR For Bangla Script", *Center for Research on Bangla Language Processing*.

[5] B.B. Chaudhuri and U. Pal, "A Complete Printed Bangla OCR System", *Pattern Recognition*, Vol. 31, No. 5, pp. 531-549, 1998.

[6] M.A Sattar, K. Mahmud, H. Arafat and A.F.M Noor Uz Zaman,"Segmenting Bangla Text for Optical Recognition", *ICCIT 2007, 10th international conference on Computer and information technology*, 27-29 Dec. 2007.

[7] R.C. Gonzalez and R.E. Woods, "Digital Image processing", second edition, *Pearson Education*, 2003.

[8] Bibhutibhushan Bandopadhyay, "Pather Panchali", *Ranjan Prakashalay*, 1929.

[9] Palit, Rajesh and Sattar, Md Abdus, "Representation of Bangla Characters in the Computer Systems". *Bangladesh Journal of Computer and Information Technology*, Vol. 7, No. 1, December 1999.

PIssGA: An ultra fast meta-heuristic approach to solve protein inference problem

Shubhra Kanti Karmaker Santu
Department of Computer Science and Engineering
Bangladesh University of Engineering and Technology
Dhaka-1000, Bangladesh
Email: kanti.shubhra006@gmail.com

Sajjadur Rahman
Department of Computer Science and Engineering
Bangladesh University of Engineering and Technology
Dhaka-1000, Bangladesh
Email: sunnysajjad@cse.buet.ac.bd

Saikat Chakraborty
Department of Computer Science and Engineering
Bangladesh University of Engineering and Technology
Dhaka-1000, Bangladesh
Email: saikatch107@gmail.com

M. Sohel Rahman
A*l*EDA Group
Department of Computer Science and Engineering
Bangladesh University of Engineering and Technology
Dhaka-1000, Bangladesh
Email: msrahman@cse.buet.ac.bd

***Abstract*—The protein inference problem represents a major challenge in proteomics research. In this paper, we propose a novel meta-heuristic approach i.e. PIssGA to infer proteins from a given peptide sequence identified from tandem mass spectrometry in shotgun proteomics. Although most of the works available in the literature mainly adopts parsimonious pressure i.e. prefer least number of inferred proteins, our model is neither strictly parsimonious nor strictly optimistic, rather a tunable model that provides the flexibility to infer proteins either parsimoniously or optimistically or somewhere in between. Another important feature of PIssGA is its efficiency in search that makes the approach ultra first. We used Sigma49 dataset to test our method and found the proposed algorithm extremely fast and moderately accurate.**

I. Introduction

Protein Inference is an intermediate step of the protein identification problem [1]. The protein inference problem has gained immense attention in the field of medical science, pharmacy, molecular biology, bio-informatics and computational data analysis. It is concerned with the identification of proteins present in a given cell or tissue which is an explicit goal of proteomics research. Study of proteins can provide information that is not readily available from genomic sequences. One of the most promising developments to come from the study of proteins has been the identification of potential new drugs for the treatment of disease. This relies on genome and proteome information to identify proteins associated with a disease, which computer software can then use as targets for new drugs. The better the accuracy of protein inferencing the greater is its impact on proteomics research.

Given a biological cell or tissue, the problem of determining which proteins are present in the sample is referred to as the protein inference problem. Now, the protein identification process may be divided into three sequential steps: *peptide identification*, *protein inference* and *result evaluation*. Firstly, a given biological sample is analyzed through a sequence of experimental and computational procedure [2], [3]. Subsequently, MS/MS spectra obtained from tandem mass spectrometry [4] is typically used to separate, fragment and finally identify [5][6] peptides from the sample. After the peptides have been identified, it is necessary to validate the identification process so that further steps do not suffer from noisy inputs. To this end, the difficulty lies in to the lack of a one to one mapping between the proteins and the peptides. Additionally, even if the identified set of peptides are reliable, it does not ensure that a reliable list of proteins can be assembled from these peptides. Two related terminologies that are commonly encountered in the protein inference problem are: one-hit wonders and degenerate peptides [7]. A protein is said to be a one-hit wonder if it generates only a single peptide when digested. It might seem that a one-hit wonder is easy to infer; however, in reality, that is not the case often. The presence of a false positive is highly likely in practice and the claim of a particular peptide being present is not absolute. Thus, protein inference faces the problem of noisy data. On the other hand, if a peptide is common between multiple proteins then that peptide is called a degenerate peptide. Thus, in the presence of a degenerate peptide, it is always difficult to deduce which protein or protein family actually generated the peptide. As a result, the protein inference problem [8], often has multiple solutions and can be computationally intractable.

Three types of models are used in the protein inference domain: statistical model (ensures protein inferencing with high accuracy), parsimonious model (assumes only a small subset of proteins should be sufficient to explain all identified peptides), and optimistic model (returns all the proteins or protein families that have some potential). In this paper, we propose a steady state genetic algorithm based approach to identify proteins from a given peptide sequence obtained from mass spectrometry. The contributions of this paper are as follows:

- Our approach provides a very fast and efficient heuristic

search strategy to infer proteins with reasonable accuracy and precision.

- Our approach provides the flexibility to infer proteins either parsimoniously or optimistically or somewhere between the two, based on some tuning parameter as discussed in Section III.

The rest of the paper is organized as follows: Section II briefly describes the related works. Section III discusses the algorithm in detail and its various features. Section IV shows the experimental setup and results. Finally, Section V draws the conclusion and gives possible future directions.

II. Related Works

Over the years, various methods have been tried and tested to conduct the protein inference process. Nesvizhskii et.al. first addressed this challenge using a probabilistic model [10], but different problem formulations and new solutions have been proposed as well [9], [11], [12]. A combinatorial approach to the protein inference problem that incorporates the concept of peptide detectability, i.e., the probability of a peptide to be detected (identified) in a standard proteomics experiment, with the goal of finding the set of proteins with the minimal number of missed peptides is presented in [11]. In the other combinatorial formulation [12], the parsimony condition is chosen albeit without any proper theoretical justification. Furthermore, parsimonious formulations often lead to the minimum cover set problem, which is NP-hard. Thus, greedy [11] or graph-pruning strategies [11] address the protein inference problem without any performance guarantee. In another approach [13], the protein inference is addressed by proposing two novel Bayesian models (Basic and Advanced) that take as input a set of identified peptides from any peptide search engine, and attempt to find a most likely set of proteins from which those identified peptides could have originated. The basic model assumes that all identified peptides are correct, whereas the advanced model also accepts the probability of each peptide to be present in the sample. The evolutionary identification approach [14] tries to find the entire sequence of a protein, even in the case of variants or unknown proteins. To accomplish that, different peptides that compose a given protein are identified as follows by first, finding their mass with a MS spectrum and then, from their mass, finding their sequences with MS/MS spectra. If a peptide can be identified correctly, the complete protein can be easily inferred.

III. Our Approach

We propose a new Meta-heuristic approach to solve the Protein Inference problem. Our approach, namely, PIssGA (The Protein Inference using Steady State Genetic Algorithm, pronounced Pizza), uses steady state genetic algorithm to solve the protein inference problem. We applied the steady state genetic algorithm for this purpose. The motivation behind choosing the steady state version of genetic algorithm is due to the exploitative nature of the protein inference problem. As with any of the genetic algorithm approaches, after initialization, PIssGA also iterates through the fitness evaluation, parent selection, breeding and survival selection procedures. The summary of the procedure is presented in Algorithm 1. The details of the problem specific procedures are presented in the later subsections.

Algorithm 1 Meta-heuristic Approach to Protein Inference

Input: S, Set of identified peptides from shotgun proteomics
Output: Q, Set of inferred protein
Method:

$P_0 \leftarrow$ Initial Population with m random candidate solutions
$n \leftarrow 0$;
while Better solution can be found **do**
  Evaluate fitness of each individual i in P_n
  $Parent_1 \leftarrow Selection(P_n)$
  $Parent_2 \leftarrow Selection(P_n)$
  $C1, C2 \leftarrow Crossover(Parent1, Parent2)$
  $C1 \leftarrow Mutate(Parent1)$
  $C2 \leftarrow Mutate(Parent2)$
  $n \leftarrow n + 1$;
  $P_n \leftarrow$ best m individuals from $(P \cup C1 \cup C2)$
end while
Output individual Q from P_n having the best fitness value

A. Representation of the candidate solution

Each candidate solution infers the proteins that are present in the sample. The simplest representation is to consider all the proteins and encode it in a vector/genome as follows: if the protein is present, the corresponding gene has value 1 and if the protein is absent, the corresponding gene has value 0. Thus a string of 0 and 1's will form an individual. Figure 1 gives an example representation which assumes five potential proteins. In Figure 1, the candidate solution represents that the testing biological sample contains proteins Pr_2, Pr_3 and Pr_5. Recall that the main database may contain many more proteins those are not potential with respect to the peptide sequences obtained from MS and MS/MS spectrum analysis [15] and thus can be discarded from the candidate solution representation as they can never appear. This point is further clarified in Section III-B.

Pr_1	Pr_2	Pr_3	Pr_4	Pr_5	Pr_6
0	1	1	0	1	0

Fig. 1. Representation of the candidate solution

B. Initial Population Generation

Given a protein, its peptide sequence is well known and available in many protein databases. Thus, by constructing an inverted index, the possible potential proteins can easily be identified given the set of identified peptides through shotgun proteomics. Thus, we need to consider only the potential proteins, not all of them, which significantly reduces the candidate solution size as well as computational complexity.

For the initial population generation, let use consider the mapping of protein to its peptide sequence. An hypothetical

mapping is demonstrated below, where Pr_n is a protein and the symbols at the right of the arrow sign actually represents its hypothetical peptide sequence. This mapping is known beforehand from some protein databases.

$$Pr_1 \rightarrow GKEUTR$$
$$Pr_2 \rightarrow LYARE$$
$$Pr_3 \rightarrow LVGARTHNB$$
$$\ldots\ldots\ldots\ldots\ldots\ldots$$
$$\ldots\ldots\ldots\ldots\ldots\ldots$$
$$Pr_n \rightarrow WHBAFGSTHJSYB$$

Now, after the MS and MS/MS spectra have been analyzed through tandem mass spectrometry, one can determine the peptides generated from the parent proteins of the test sample. Though this peptide identification process is not perfect, it can be accomplished with reasonable accuracy. Now, if the aforementioned idea of an inverted index is applied, one can easily determine the potential parent proteins expected to be present in the given sample as follows. All we have to do is to take a peptide generated from the shotgun proteomics and then search the inverted index to find the potential parent proteins that might have generated the peptide. More elaborately, if protein Pr_1 generates a peptide sequence which contains the peptide G, then the former is obviously a potential parent of the latter. In this way, we find all the potential proteins in the sample, form a set of potential proteins removing any repetitions and discard the rest of the proteins as they are of no interest for this particular test case.

As the number of proteins present in the test sample cannot exceed the cardinality of the potential protein set, number of genes in an individual is kept equal to the cardinality of the latter. This helps to reduce the size of the candidate solution and reduce the search space. After the number of potential proteins in an individual and also the individual potential proteins have been identified, we are ready to generate the initial population. We generate the individual of the initial population by just generating a random string of 0's and 1's. A sample initial population with 4 individuals is presented in Figure 2. To understand this figure properly, let, there are n proteins (from Pr_1 to Pr_n) in a protein database, some of which are selected (by the inverted index procedure mentioned previously) as potential proteins that might be present in a particular sample. Let the potential protein set is $\{Pr_3, Pr_4, Pr_9, Pr_{11}, Pr_{14}\}$. So, the candidate solution representation considers these potential proteins only discarding the rest. As shown in figure 2, the first candidate solution infers proteins P_4 and P_9 to be present, whereas the third individual infers proteins P_9 and P_{11}.

C. Breeding operators

The main two breeding operators used in PIssGA are the Recombination and Mutation. Let us first consider the recombination operator. As the candidate solution is encoded as a boolean valued vector and the sequence of genes does not have any special significance, uniform crossover [18] is applied. This ensures that every gene has equal probability of being swapped. This swapping probability is denoted by γ which is a user defined parameter. Figure 3 demonstrates this process where uniform crossover swaps the genes corresponding to Pr_1, Pr_3 and Pr_6.

Pr_3	Pr_4	Pr_9	Pr_{11}	Pr_{14}
0	1	1	0	0
1	1	0	0	1
0	0	1	1	0
1	0	1	0	1

Fig. 2. A sample initial population

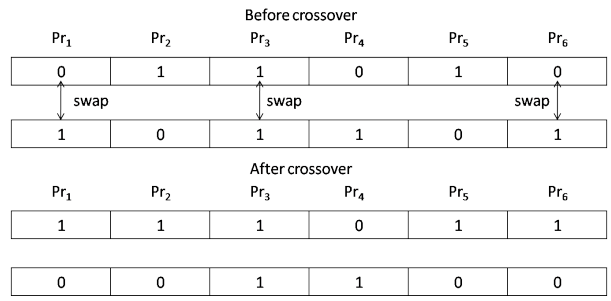


Fig. 3. Uniform Crossover

For mutation, bit-flip mutation operator is applied. Here, we randomly select a gene from an individual and with a very small probability μ, we flip the value of that gene. Figure 4 demonstrates this process where the mutation operator flips the value of Pr_3 from 1 to 0.

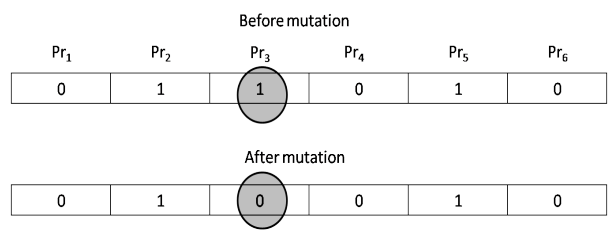


Fig. 4. Bit flip mutation operator

D. Selection procedure

Parent selection for breeding in PIssGA is performed using the roulette wheel selection mechanism, also known as proportionate reproduction [19]. It selects individuals for crossover and mutation operations. The individuals of a population are selected for genetic operations with a probability proportional to their fitness value defined by Eq. 1 to be discussed in the following section. So, the fitter individuals have a higher probability of being selected.

The survival selection follows the exploitative approach. PIssGA only replaces a parent individual only when the child has a greater fitness.

E. Fitness evaluation

While evolving the candidate solutions, the fitness function should consider the following issues:

- One issue lies in the choice of whether we should consider the minimum or maximum number of proteins that cover the test peptide sequence or whether to keep both the options and make it adaptive. To this end, PIssGA takes an adaptive approach (treating it as a user defined parameter). This is one of the unique features of PIssGA which gives user the control over these options.
- Percentage of test peptides covered by the proteins inferred by a candidate solution
- Another issue is how much the hypothetical peptide set constructed from the inferred proteins of an individual is similar to the test peptide sequence. The more similar they are, the higher is the probability of the inferred protein sets to be correct. Also important is the issue of how many redundant peptides does this hypothetical peptide set contains.
- The protein inference procedure is not deterministic due to the presence of both one-hit wonders and degenerate peptides. So, an exact solution may not be available and approximate solutions may be only possible option.

For the ease of notation, let us assume that we have the following functions:

- **N**(c)**:** Returns the number of 1's in an individual c.
- **C**(c)**:** This is the coverage function which returns how much the hypothetical peptide set constructed from the inferred proteins of individual c cover the test peptide sequence.
- **R**(c)**:** This is the redundancy function which returns how much the hypothetical peptide set resulting from inferred proteins of individual c contains redundant proteins with respect to the test peptide sequence.

For a better understanding of the Coverage and the Redundancy measure, let us consider Figure 5. Here, Coverage means the intersection of the inferred hypothetical peptide set (A) and the test peptide set (B), i.e., $A \cap B$. Redundancy contains the peptides in the inferred hypothetical peptide set that are not part of the test peptide set, i.e., $A - B$.

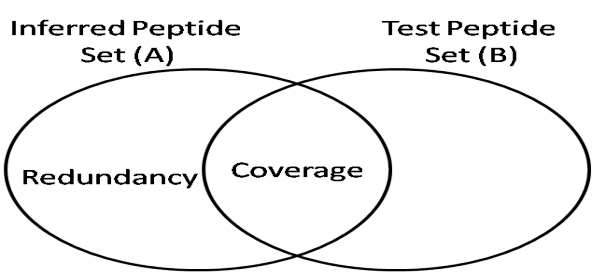


Fig. 5. Demonstration of coverage and redundancy

Now, the fitness function can be expressed in the following manner:

$$F(c) = [(C(c))^{\alpha} - (R(c))^{\beta}] * (N(c))^{\epsilon} * \left(\frac{1}{N(c)}\right)^{1-\epsilon} \quad (1)$$

The fitness function defined by equation 1 tries to maximize the coverage and minimize the redundancy simultaneously. Here, α and β are two weighting exponent fractions. These two constants are varied for 0 to 1 and adjusted by experiment. ϵ is a user defined parameter, which indicates the user preference for whether to take maximum possible number of proteins or minimum possible number of proteins as the desired output or something in between the two. Note that, the value of ϵ is varied between 0 and 1: here, 0 (1) means that the user is preferring to take minimum (maximum) possible number of proteins, whereas extreme end 1 means the user is preferring to take maximum possible number of proteins. ϵ gives PIssGA the flexibility to prefer either high precision or high recall or a mid way around. Notably such adaptive options were not available in the previous Bayesian approaches. The Naive Bayes approach with multiplication of probabilities would always prefer minimum number of proteins and thus is biased towards higher precision while obtaining poor recall.

IV. Experiments

A. Dataset

We used Sigma49 dataset collected from [21] to test our algorithm. Sigma49 sample was prepared by mixing 49 human proteins, among which 44 proteins contain at least one peptide that can be identified by shotgun proteomics. In addition, 9 keratin proteins and 4 other proteins are categorized as the keratin contamination and bonus proteins, respectively, and are believed to be present in the sample due to contamination.

Sigma49 provides 3 replicates (the same experiment run thrice) of Peptide Prophet results which contains the peptide search results based on the MS/MS spectra obtained from tandem mass spectrometry. It also contains the probabilities of the predicted peptides to be correct. These search results were used as the inputs to our approach which starts from these identified peptides and finally infers the proteins.

B. Experimental Setup

We implemented our PIssGA on a core i-3 intel microprocessor, 2 GB RAM machine using JAVA. Peptide Prophet result of sigma49 dataset was parsed to peak out identified peptides. At first probable proteins predicted by Peptide Prophet Result are given a peptide count as per Peptide Prophet result. All the peptides belonging to proteins having peptide count 2 or more are directly taken as identified peptides. For those peptides belonging to proteins with peptide count 1, predicted probability is checked and filtered with minimum probability 0.97.

C. Performance measures

Following the literature, we use the performance measures True Positive, False Positive, False Negative, precision, recall and F-measure [20]. They are briefly discussed below in reference with the confusion matrix [20] shown in figure 6.

Precision: It refers to the portion of correctly inferred proteins within the test peptide set that are inferred to be present by the protein inference algorithm.

$$precision = \frac{TP}{TP + FP} \quad (2)$$

		True class	
		P	N
Hypothesised class	Y	True Positive (TP)	False Positive (FP)
	N	False Negative (FN)	True Negative (TN)

Fig. 6. Confusion matrix

	MMP	PP	BB	BBA	ABP	ABPA	PIssGA
TP	39	41.5	39	37	35	43	40.5
FP	6	7.5	16	6	4	22	8
FN	5	2.5	5	7	9	1	3.5
PR	0.87	0.85	0.71	0.86	0.9	0.66	0.84
RC	0.89	0.94	0.89	0.84	0.8	0.98	0.92
F	0.88	0.89	0.79	0.85	0.84	0.79	0.88

TABLE I
COMPARISON OF RESULTS

Recall: It is also called the true positive rate. It is the portion of peptides within the test peptide set that are correctly inferred by the protein inference algorithm.

$$recall = \frac{TP}{P} \tag{3}$$

F-measure: It is the harmonic mean of precision and recall.

$$f_measure = \frac{2}{\frac{1}{precision} + \frac{1}{recall}} \tag{4}$$

D. Parameters

The parameters used in this work are uniform crossover probability γ, mutation probability μ and the special parameters ϵ, α and β. The value of γ, μ, α and β were set at 0.5, 0.1, 0.25 and 0.5 respectively to get the best results. That is, during crossover, each bit was swapped with a probability of 0.5, whereas, the bit flip mutation flipped the bit with probability 0.1. α was set to a lower value than β meaning that we emphasized coverage more than redundancy, as both coverage and redundancy are within the scale of 0 to 1. On the other hand, ϵ was varied between 0 and 1 to observe the effects of parsimonious pressure and optimistic pressure.

E. The evolution of candidate solutions

The fitness as well as the precision, recall, F-measure associated with the candidate solutions evolve and gradually get better as the number of iterations increase. Figure 7, 8, 9 and 10 shows the dynamics of the associated fitness, precision, recall and F-measure of the best individual evolved along with the number of iterations of the evolutionary process. It is very clear from these figures that, as the number of iterations increases, the solutions approach towards higher fitness, precision, recall and F-measure. At the initial stages, the changes are more rapid than in the final stages where the candidate solutions have nearly converged. This justifies the validity of the proposed evolutionary approach.

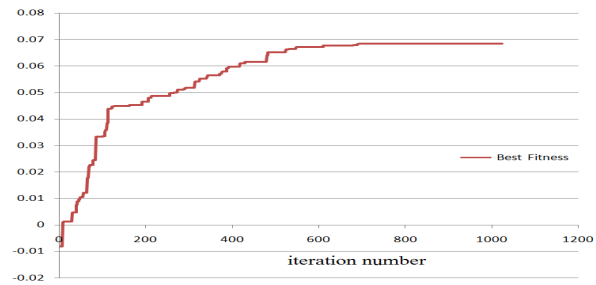


Fig. 7. Evolution of fitness

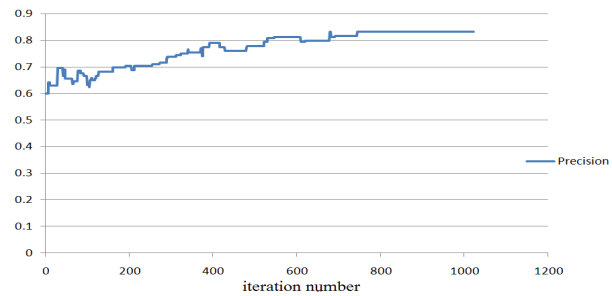


Fig. 8. Evolution of precision

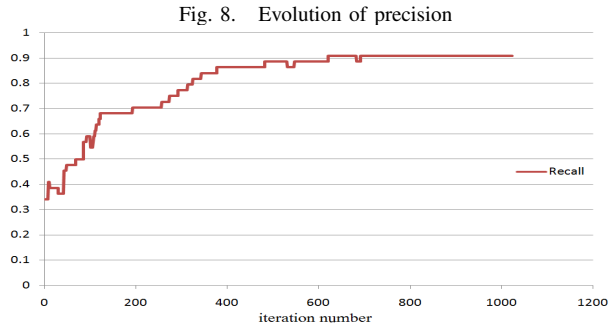


Fig. 9. Evolution of recall

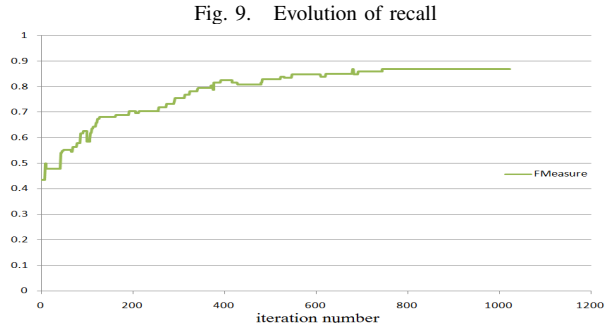


Fig. 10. Evolution of F-measure

F. Results

Protein inference results on the Sigma49 dataset using minimum missed peptide approach (MMP)[17], ProteinProphet (PP) [16], basic Bayesian model (BB)[13], basic Bayesian model with detectability adjustment (BBA)[13], advanced Bayesian model using raw PeptideProphet probabilities (ABP)[13], ABP after detectability adjustment (ABPA)[13] and our proposed Meta-heuristic approach (PIssGA). All results are evaluated based on the true positive (TP), false positive (FP) and false negative (FN), precision (PR), recall

(RC) and F-measure (F) as discussed in Section IV-C. Table I shows the summary result.

In our experimental setup we tried to infer 44 proteins of Sigma49 as the rest five proteins did not contain a single peptide that could be identified by the Peptide Prophet search. The results show that we have a better Precision than BB and APBA and comparable Precision with MMP, PP, BBA, ABP. In terms of Recall, PIssGA outperforms majority of the existing methods (MMP, BB, BBA, ABP). Again in terms of F-measure PIssGA exhibits better result than BB, BBA, ABP, ABPA and comparable result to MMP and PP. It should be noted that although the results obtained by PIssGA does not outperform all the other approaches by a great margin, it achieves very competitive results drastically faster than the other approaches. More elaboration about the running time is presented in the following paragraph.

In brute force method of protein inference it takes $O(2^m)$ time if candidate protein size is m. But PIssGA gives near optimal results spending much less time. A box plot of the running time for both PIssGA and brute force bayesian approaches is shown in Figure 11 in case Sigma49 dataset. The picture shows clearly that PIssGA is drastically faster than the previous approaches. The huge reduction in the computational time offers biologists to get very quick results for elementary evaluation of protein samples.

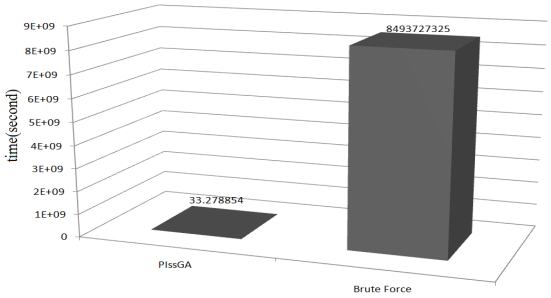


Fig. 11. Comparison of time between PIssGA and Brute force approach

V. Conclusion

In this paper we have presented PIssGA to solve the protein inference problem and measured its performance on Sigma49 dataset. The experimental results show that, our algorithm outperforms several of the existing protein inference methods. Given a peptide sequence, our algorithm is drastically faster in terms of identifying proteins than the existing approaches. PIssGA provides a novel meta-heuristics approach for protein inference which enables a new research direction. The results obtained so far are promising and further investigation with other variants of genetic algorithm may yield better results.

A possible future work may be to introduce the detectabilities [22] proposed by Li et.al. into our approach which is likely to increase the accuracy of the predictor. Another direction can be to formulate the problem as a multi-objective optimization of coverage maximization and redundancy minimization. Furthermore, hybrid approaches can be proposed that combines both the features of meta-heuristic and bayesian approaches.

References

[1] Protein inference: a review. Huang, Ting and Wang, Jingjing and Yu, Weichuan and He, Zengyou - Briefings in Bioinformatics (10.1093/bib/bbs004), 2012 - leeds-faculty.colorado.edu

[2] Aebersold, R., Mann, M. Mass spectrometry-based proteomics. Nature 422, 198207 (2003)

[3] Kislinger, T., Emili, A. Multidimensional protein identification technology: current status and future prospects. Expert Rev. Proteomics 2(1), 2739 (2005)

[4] Alexey I. Nesvizhskii[1], Ruedi Aebersold[2]. Interpretation of Shotgun Proteomics Data: The Protein Inference Problem. [1]Institute for System s Biology,1441 N 34th Street, Seattle WA 98103 USA. [2]ETHZ Federal Institute of Technology, CH-8092 Zurich, Switzerland

[5] Marcotte, E.M. How do shotgun proteomics algorithms identify proteins?. Nat. Biotechnol. 25(7), 755757 (2007)

[6] Nesvizhskii, A.I. Protein identification by tandem mass spectrometry and sequence database searching. Methods Mol Biol 367, 87119 (2007)

[7] T Huang, J Wang, W Yu Protein inference: a review Briefings in Bioinformatics, 2012 - Oxford Univ Press

[8] Nesvizhskii, A.I., Aebersold, R. Interpretation of shotgun proteomic data: the protein inference problem. Mol Cell Proteomics 4(10), 14191440 (2005)

[9] Nesvizhskii, A.I., Aebersold, R.: Interpretation of shotgun proteomic data: the protein inference problem. Mol Cell Proteomics 4(10), 14191440 (2005)

[10] Nesvizhskii, A.I.,Keller, A., Kolker, E., Aebersold, R.: A statistical model for identifying proteins by tandem mass spectrometry. Anal Chem 75(17), 46464658 (2003)

[11] Alves, P., Arnold, R.J., Novotny, M.V., Radivojac, P., Reilly, J.P.,Tang, H.: Advancement in protein inference from shotgun proteomics using peptide detectability. In: PSB 2007: Pacific Symposium on Biocomputing, pp. 409420. World Scientific, Singapore (2007)

[12] Zhang, B., Chambers, M.C., Tabb, D.L.: Proteomic Parsimony through Bipartite Graph Analysis Improves Accuracy and Transparency. J Proteome Res. 6(9), 35493557 (2007)

[13] Yong Fuga Li, Randy J. Arnold, Yixue Li, Predrag Radivojac, Quanhu Sheng, and Haixu Tang1: A Bayesian Approach to Protein Inference Problem in Shotgun Proteomics. RECOMB 2008, LNBI 4955, pp. 167180, 2008.

[14] Jean-Charles Boisson, Laetitia Jourdan and El-Ghazali Talbi, Christian Rolando. A Prelimainary Work on Evolutionary Identification of Protein Variants and New Proteins on Grids. In Proceedings of the 20th International Conference on Advanced Information Networking and Applications 2006.

[15] Alexey I. Nesvizhskii[1], Ruedi Aebersold[2]. Interpretation of Shotgun Proteomics Data: The Protein Inference Problem. [1]Institute for System s Biology,1441 N 34th Street, Seattle WA 98103 USA. [2]ETHZ Federal Institute of Technology, CH-8092 Zurich, Switzerland

[16] Nesvizhskii, A.I., Keller, A., Kolker, E., Aebersold, R. A statistical model for identifying proteins by tandem mass spectrometry. Anal Chem 75(17), 46464658(2003)

[17] Alves, P., Arnold, R.J., Novotny, M.V., Radivojac, P., Reilly, J.P., Tang, H. Advancement in protein inference from shotgun proteomics using peptide detectability. In: PSB 2007: Pacific Symposium on Biocomputing, pp. 409420. World Sci-entific, Singapore (2007)

[18] Uniform Crossover in Genetic Algorithms by: Gilbert Syswerda edited by: David J. Schaffer In Proceedings of the Third International Conference on Genetic Algorithms (1989), pp. 2-9

[19] D. E. Goldberg and K. Deb. A comparative analysis of selection schemes used in genetic algorithms. InFoundations of Genetic Algorithms, pages 69-73. Morgan Kaufmann, 1991.

[20] Tom Fawcett . An introduction to ROC analysis

[21] http://darwin.informatics.indiana.edu/yonli/publicdata/

[22] Yong Fuga Li, Randy J. Arnold, Haixu Tang, and Predrag Radivojac (2010). The importance of peptide detectability for protein identification, quantification, and experiment design in MS/MS proteomics. Journal of Proteome Research. 9(12): 62886297

Categorization of Copies to Improve the Performance of Routing Schemes in Delay Tolerant Networks

Tanmoy Naha, Shaikat Barua, Shahid Md. Asif Iqbal, Alok Kumar Chowdhury, Kingshuk Dhar
Department of Computer Science & Engineering
Premier University, Chittagong
Bangladesh.
{nahapu211, shaikat213}@gmail.com, {asifcsep, alokchy04, kingshuk2006}@yahoo.com

Abstract— **Optimizing the three most important network performance parameters i.e. Delivery probability, Overhead ratio, and Delay is challenging in Delay Tolerant Networks (DTN) due to high occurrences of network partitioning in these networks. Owing to the partitioning nature efficient routing or dissemination of messages is important in DTN. In this paper, we focused to improve the spray phase of Binary Spray and Wait routing scheme, an efficient routing scheme in DTN, by dividing the number of copies of a message in different categories. The proposed scheme provides higher delivery probability and lower delay for messages while keeping the overhead ratio low. Here, an adaptive spraying scheme is used combining the regular binary spraying and performance based spraying of a message in each node. Thus, each message is forwarded starting with the binary followed by an adaptive spraying technique depending on the time to live (TTL) value of that message. We also formulate an equation to permit a node to calculate a switch value for each message it carries. Later, the switch value is compared to the TTL of the message to toggle to the wait phase for that particular message instead of the number of copies of the message reaching to one. Simulation results show that our adaptive scheme has better delivery and delay performance than that of other schemes.**

Keywords: Adaptive Spraying, Switch Value, Regular Copy, Performance Copy, Visit List, Delivery Predictability

I. INTRODUCTION

Nodes in a Delay Tolerant Networks usually form group or partition in the network because of intermittent connectivity and unstable routing paths. Thus, DTN is a kind of mobile ad hoc network where the partitioning of nodes is common or frequent and sometime the network topology is so sparse that each partition contains only one node. This particular nature of these networks allows nodes to communicate with each other without establishing complete end-to-end paths or connectivity between them. Occasionally, nodes may find end-to-end paths or direct connection between them but the dominant feature i.e. network partitioning compels these paths or connections to be broken down or change soon after it has been discovered. A DTN which is covering a fairly sparse topology can be considered as a set of disconnected, time-varying clusters of nodes [1]. However, many real scenarios like Interplanetary Internet, Deep Space Networks, Sensor-Based Networks, Military Networks, Inhabitant or Wildlife Tracking System, Terrestrial Wireless Networks, Satellite Networks, Underwater Acoustic Networks, Nomadic Communities Networks [1]-[2] are ideal for application or deployment of DTN. The factors or reasons behind DTNs suffering extremely unsteady paths include high node mobility, low node density, intermittent power from energy management schemes, environmental interference and obstruction, short range radio and malicious attacks [2] etc.

Intermittent connectivity and lack of end-to-end path between nodes cause well-known ad hoc routing schemes to fail in DTN. Thus, routing is one of the real or non-trivial problems in DTNs that encouraged researchers to be attentive here. Routing schemes proposed for DTN so far are specially developed to overcome the problems and challenges of DTNs and all of them are based on "store-carry-forward" mechanism. Instead of using packet switching or circuit switching, which is not possible for lack of end-to-end paths, nodes in DTN mainly use message switching. A message is buffered or carried by the source or a relay node for a long time until the destination is reached or an intermediate node is seen or the TTL of the message expires. Routing schemes that are proposed to work with DTN can be divided into two broad groups: *Forwarding Based* and *Replication Based*. Forwarding based routing schemes make use of a single copy of a message to deliver it to the destination. This group of routing schemes mainly works with the assumption of network connectivity and environmental knowledge to take routing decisions [2]. However, forwarding based routing schemes perform poorly when the environment is completely opportunistic. On the other hand replication based routing schemes take the advantage of temporal connectivity to make a sequence of independent and local forwarding decisions. Replication based routing schemes inject multiple copies of each message into the network to ensure high delivery probability. The authors in [1] considered the issue of how many replicas of a message should be injected into the network. The more the number of copies injected the better

the chance to increase the delivery ratio and decrease the latency. However, too many copies of each message may result in large communication overhead.

In this work, we propose to improve the spray phase of Binary Spray and Wait routing. The main contribution of our spray phase is as follows: first we divide the number of copies of a message into two broad categories, regular copies and performance copies. Again, we divide the performance copies into two parts, visit copies and predict copies. Whenever a node, either source or relay, comes in contact with another node it transmits copies from both categories of a message to the node if the node does not possess any copy of that particular message.

The nodes mainly distribute regular copies first followed by the forwarding of performance copies. Like [3], we also calculate a *switch* value for a message in each node and use it to permit the node to enter into the wait phase for that particular massage. Simulation results show that our proposed scheme maintains good delivery and latency performance while keeping the communication overhead low. The rest of the paper is organized as follows. Chapter II and chapter V describe the related works and conclusion of the paper respectively. In chapter III we depict our new performance based routing scheme. Simulation setup and results analysis come in chapter IV. We also give some future research directions in chapter V as well.

II. RELATED WORKS

Though challenging at the very beginning researcher has been successfully proposing routing schemes for DTN throughout the last decade. Among this routing schemes replication based routing schemes are well known for having high delivery ratio and lower delay as they inject multiple copies of a massage into the network. However, injecting multiple or redundant copies of a message demands more network resources like buffer space, bandwidth and energy [4]. Replication based schemes are again divided into two groups flooding based schemes and quota based schemes. Epidemic[4], PRoPHET[5], MaxProp[6], RAPID[7], PREP[8] etc. are considered to be flooding based schemes as these schemes inject or replicate as many copies of a message as the network permits which creates problem during high network overhead and latency. Quota based routing scheme like Spray and Wait [1], ORWAR [9], Spray and Focus [10], and EBR [11] relay limited no of copies of a message in the network and they are known to be resource savvy or resource friendly. However injection of limited no of copies leads these schemes to suffer from lower delivery ratio. Nevertheless, existing quota based routing schemes experience comparatively lower delivery ratio even though they are the better steward of network resources.

In PRoPHET [5], nodes use delivery predictability to guide a massage to its destination. When two nodes meet, the node with the more delivery predictability of the desired destination wins the message destined for that destination. However, PRoPHET suffers from lower delivery and higher delay when the nodes have low buffer size. These also increase the average number message forwarding or relaying which may eventually lead to high overhead. Spray and Wait (SW) [1], allows a source node to initially generate L number of copies of a message. The scheme makes use of two phases; Spray phase and Wait phase. During the Spray phase, the source node forwards the L copies of a message to L nodes which do not have a copy of the message. If the destination is not among these L nodes, then the scheme switches to the Wait phase where the L nodes keep their copy of the message until the destination is reached. Binary Spray and Wait (BSW) which is similar to the SW exhibits the same spreading of a message except that each node forwards half the copies of a message to a newly seen node instead of forwarding a single copy. When a node is left with only a single copy of a message, it switches to the Wait phase and performs direct transmission when it meets the destination. However, both SW and BSW do not take into consideration the performance and blindly flood the message into the network. Moreover SW & BSW treat all the copies of a message equally or linearly which prevents intelligent spreading of the message. But, our scheme divides of discriminates the copies of a message into regular copies & performance based copies. This helps our scheme to take the advantage of both legacy schemes, e.g., SW and BSW and our new performance based spreading technique. Recently several schemes [12]-[15] have been proposed which are mainly improvement of SW and BSW. In [12], a spray parameter is defined using the current no. of copies and initial no. of copies, later a spraying technique e.g. BSW, SW, or Adaptive was selected using two limit parameters α_1 and α_2. Agents of source nodes of messages were used for faster and efficient spreading of messages in [13]. In [14], authors calculate a utility metric to spray the copies of a message into the network. Average delivery predictability has been used by a node in [15] to decide whether to spray copies of a message or not when it counters a new node. Authors in [16] worked on the Wait phase to increase the delivery utility but experiencing low latency and low overhead. However, none of these works deem to discriminate among the copies of message which is necessary to take the advantages of both legacy spraying and performance based spraying techniques.

III. PROPOSED SCHEME

We propose to discriminate among the initial no of copies generate by the source of a message during the spray phase instead of treating all the copies equally or linearly. We propose to divide the no. of copies into two broad categories namely regular copies and performance copies. The performance copies are further grouped in two parts, visit copies and predict copies. Like [3], we also permit a node to forward a copy of a massage, either regular or performance copy, to other nodes as long as the TTL of the message is higher than that of the *switch* value. Each source or relay node of a message starts with the regular copies and sprays these copies in the same way as BSW but this phenomenon stops as soon as the TTL of the message falls below the one third value of the initial TTL or the TTL falls below the *switch* value. After that, spraying of regular copies stops and nodes

enter into the wait phase for these regular copies. The spraying of performance copies started from here after and continues until the TTL falls below the *switch* values or the no. of performance copies reaches to one. Then, the nodes enter into the wait phase for the performance copies.

A. Regular Copies:

We calculate the regular copies (*rc*) of a message in the source node using eq. (1)

$$rc = tc/2 \quad \ldots\ldots (1)$$

where , *ic* = initial no. of copies and rc = no. of regular copies After that the source node sprays these messages using the same spraying technique as BSW. Intermediate or relay nodes exhibit the same spraying technique and the binary spraying continues until the TTL is higher than the one third of the initial TTL or the TTL falls below the *switch* margin.

B. Performance Copies:

Using eq. (2), we calculate the performance copies (*pc*) of each message generated at the source.

$$pc = tc - (tc/2) \quad \ldots\ldots (2)$$

Again, we divide the performance copies in two groups, (i) visit copies *(vc)* and (ii) predict copies *(prc)* using eq. (3) & (4) respectively.

$$vc = pc/2 \quad \ldots\ldots (3)$$

$$prc = pc - v \quad \ldots\ldots (4)$$

1. **Visit list:** Like [13], each node **i** keeps a list of its visited or seen nodes as follows $V_i \subseteq N$, where **N** is the set of (n-1) nodes in the network
2. **Delivery predictability:** Like PROPHET [5], we calculate the delivery predictability, P(a,b) [0,1], in each node A for each known destination B. The delivery predictability indicates the likelihood of node A delivering a message to a particular destination B. When two nodes encounter each other, we permit the nodes to exchange the P-values of their known destinations and use the information to update the delivery predictability information at both ends.

C. Adaptive Spraying:

When a node **A,** carrying a message with n_1 number of visit copies (*vc*) and n_2 number of predict copies (*prc*), i.e. a total of $p = (n_1 + n_2)$ number of performance copies for a node **C,** meets node **B,** we allow node **A** to use eq. (5) to calculate the number of performance copies to be forwarded to **B**.

$$\left.\begin{aligned} B_{vc} &= ceil\left(\frac{B_v \times vc}{A_v + B_v}\right) \\ B_{prc} &= ceil\left(\frac{P_{(B,c)} \times prc}{P_{(B,c)} + P_{(A,c)}}\right) \\ B_{pc} &= B_{vc} + B_{prc} \end{aligned}\right\} \quad \ldots\ldots (5)$$

Thus, node **A** uses equation (6) to retain its share of the *pc* copies of the message destined for C

$$\left.\begin{aligned} A_{vc} &= vc - B_{vc} \\ A_{prc} &= prc - B_{prc} \\ A_{pc} &= A_{vc} + A_{prc} \end{aligned}\right\} \quad \ldots\ldots (6)$$

A node having more than one *pc* of a message will repeatedly do the same thing until *pc* reaches to one. But, to minimize the communication overhead in this adaptive technique we tend to be little conservative when the performance of the two nodes are unequally likely. Thus, we use a *min* value when the difference between the performance, including the visit list and delivery predictability, of the two nodes in contact is uneven. Therefore, node A and B will select the lower of the *min* value and the *pc* value. This conservative approach helps our scheme to distribute the messages evenly so that the communication overhead can be controlled. However, the *min* value depends on the number of nodes in the network, initial number of copies of a message, buffer size, etc. and we adjust the value of *min* during simulation to obtain the optimum performance

D. Wait Phase:

We have three conditions or three stages for a node to enter into the Wait phase throughout the network life time.

1. If the TTL of a message falls below the one third of the initial TTL the nodes enter into the Wait phase for the regular copies of the message.
2. If the TTL of a message falls below the *switch* value, as in [3], the nodes enter into the wait phase for both categories of copies.
3. If the number of performance copies reaches to one.

IV. PERFORMANCE EVALUATION

We implement our scheme in ONE (Opportunistic Network Environment) [17] simulator developed by Helsinki University of Technology. To validate our scheme we compare the performance our scheme with that of Spray and Wait routing, Binary Spray and Wait Routing and Adaptive Spray and Wait routing scheme [3]. We use a typical scenario consists of a 4500X3400 m^2 area of Helsinki city. Fig. 1 shows a snapshot of our simulation scenarios. We considered 126 nodes divided into 6 groups. Group 1 and Group 3 are pedestrians group and these two groups consist of 40 nodes separately.

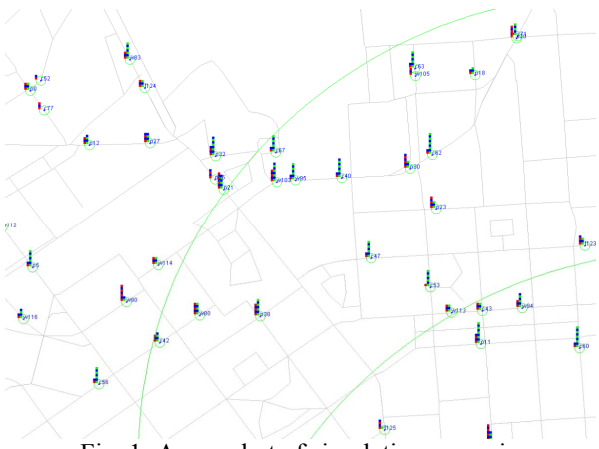


Fig. 1: A snapshot of simulation scenario

Group 2 is car group and it also consists of 40 nodes. Group 4, Group 5 and Group 6 are tram groups and these three groups consist of 2 nodes separately. Pedestrians move with speeds of 0.5 – 1.5 m/s and waiting time of 0 – 120 seconds. Cars move with speeds of 10 – 50 km/h and waiting time of 0 – 120 seconds. Trams move with speeds of 7 – 10 km/h and waiting time of 10 – 30 seconds at every station. We use two types of devices in the experiments; one is the bluetooth devices having transmission range of 10 m and transmission speed of 2 Mbps, other is the 802.11b WLAN devices with transmission range of 30 m and transmission speed of 4.5 Mbps. Group 1 through Group 3 have a buffer size of 5 MB and for the movement of those groups we use the Shortest Path Map-Based movement model. Group 4 through Group 6 have a buffer size of 50 MB and for the movement of those groups we use the Route Map-Based movement model.

We run the simulation for 12 hours. In the beginning of the simulations, we allow a warm up period of 1000 seconds. In the simulation, we vary the initial number of copies of a message generated at a source node from 2 to 15 to compare the routing schemes in different network load. In order to asses the impact of message size and node's buffer size, we also vary the message size from 1KB to 1.4MB and buffer size from 10MB to 70MB. In all the scenarios we keep the TTL of a massage to 300 and we set the value of *min* to 2 (two). We have used three performance parameters to compare the performances of the schemes.

Delivery Ratio: It is the fraction of the generated messages that are correctly delivered to the final destination within a given time period.

Delay: It is the difference between the time a message is generated and it is delivered to the destination.

Overhead Ratio: It is the ratio of the number of messages relayed and the number of messages delivered to the destination.

Fig. 2, 3 and 4 report the Delivery Ratio, Delay and Overhead ratio by varying the number of copies of a message but setting the message size to 500KB-1MB. In these figures we labeled our scheme as "Proposed".

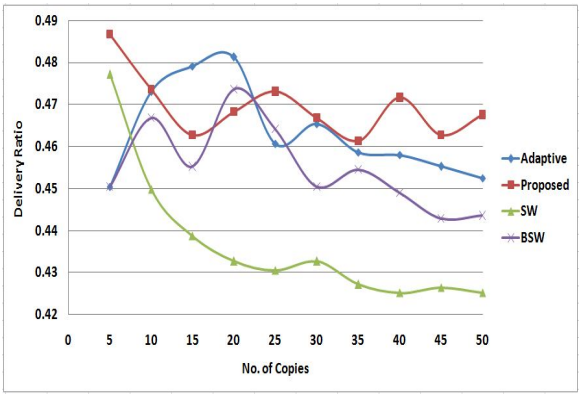


Fig. 2: Delivery Ratio: varying no. of copies (TTL 300)

Our scheme outperforms the previous routing schemes in terms of delivery ratio and delay performance lower overhead ratio than that of others. Since, we categorized the copies of a message into regular and performance copies our scheme takes the advantage of both legacy and adaptive spraying. We permitted the nodes to spray the regular copies first which helps in faster and wide spreading of messages. The performance based copies are sprayed after the TTL of a message falls below the one third value of the initial TTL which enriches the visiting list of a node and updates the delivery predictabilities of a node for different destinations. Thus, our scheme experiences good delivery and delay performances with lower overhead ratio.

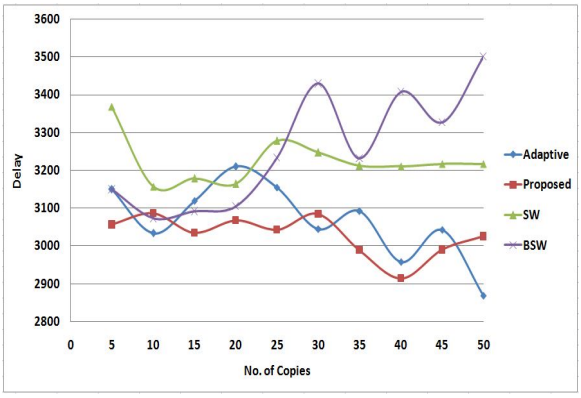


Fig. 3: Delay: varying no. of copies (TTL 300)

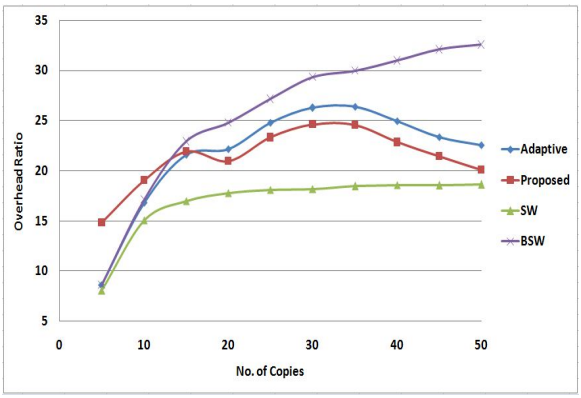


Fig. 4: Overhead Ratio: varying no. of copies (TTL 300)

We also tested the Delivery ratio (Fig. 5, 8), Delay (Fig. 6, 9) and Overhead (Fig. 7, 10 by varying both the buffer size and message size individually. The results show that our scheme performed well in terms of delivery ration, delay and overhead ration due to the reasons mentioned above.

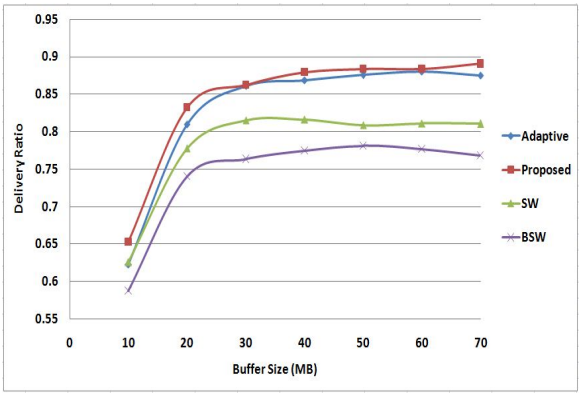


Fig. 5: Delivery Ratio: varying buffer size (TTL 300)

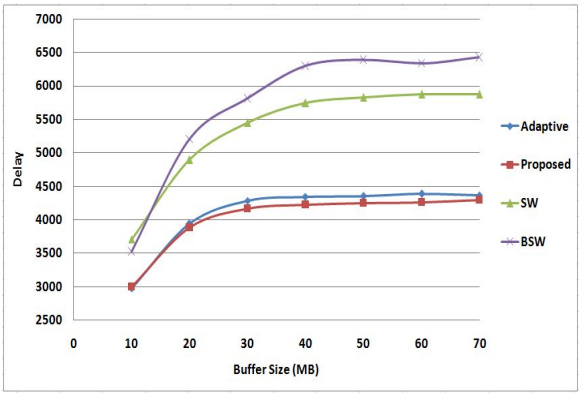


Fig. 6: Delay: varying buffer size (TTL 300)

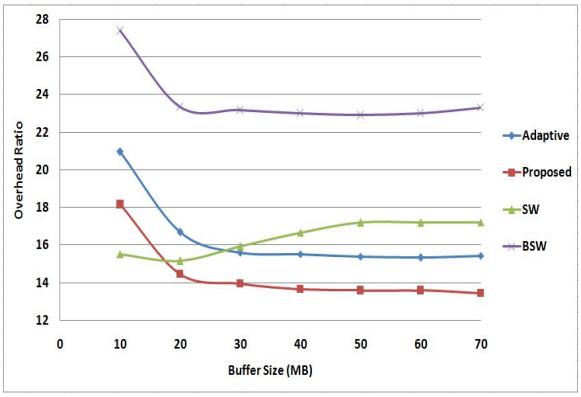


Fig. 7: Overhead Ratio: varying buffer size (TTL 300)

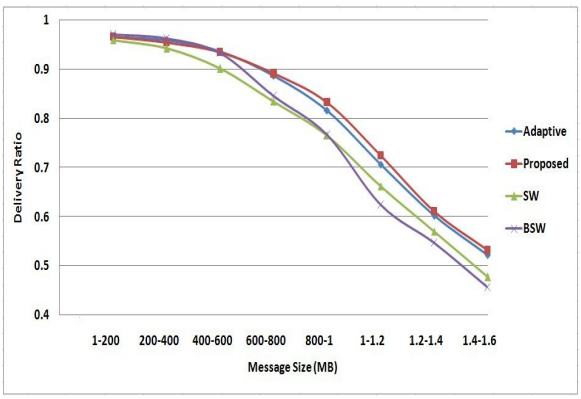


Fig. 8: Delivery Ratio: varying message size (TTL 300)

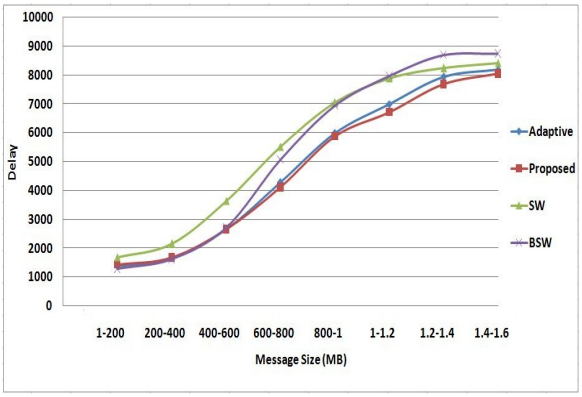


Fig. 9: Delay: varying message size (TTL 300)

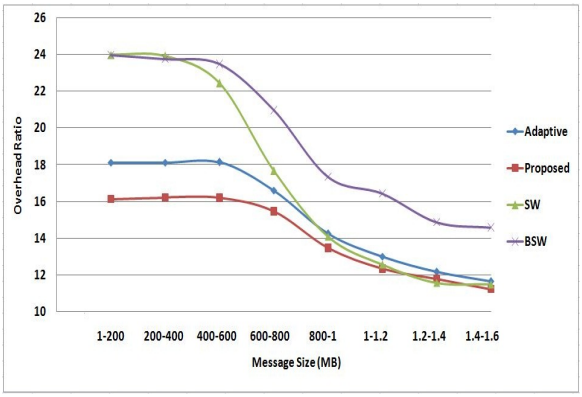


Fig. 10: Overhead Ratio: varying message size (TTL 300)

V. CONCLUSSIION

Delivering messages efficiently in Delay Tolerant Networks is challenging due to frequent connection interruption and network partitioning. Here, we proposed a performance based adaptive spraying technique that categorizes the no. of copies of a message in two groups, regular copies and performance copies, and the first category which are spreaded in the network using the legacy or blind spraying. The second category, i.e. performance copies are forwarded based on visit list and delivery predictability. The proposed scheme shows better delivery and delay performance with less communication overhead. Simulation results also stand behind our clam of achieving the objective. Our next step is to increase the efficiency by comparing with other popular Quota based routing schemes as well as Replication based routing schemes by varying our adaptive spraying technique. Finally, optimizing our scheme to achieve obstruction less network can be another direction of our future work.

REFERENCES

[1] T. Spyropoulos, K. Psounis, and C. S. Raghavendra, "Spray and Wait: An Efficient Routing Protocol for Intermittently Connected Mobile Networks," *Proceedings of the 2005 ACM SIGCOMM Workshop on Delay-tolerant Networking*, USA, pp. 252-259, August 2005.

[2] Vinod Venkataraman, Hrishikesh Bhatt Acharya, Harsh Shah, Simon Lam, "*Delay Tolerant Networking - A*

Tutorial", Department of Computer Sciences, The University of Texas at Austin, Austin, TX 78712-0233.

[3] S.M.A. Iqbal, and A.K. Chowdhury, "Adaptation of Spray Phase to Improve the Binary Spray and Wait Routing in Delay Tolerant Networks", *International Conference on Computer and Information Technology (ICCIT-2012)*, Chittagong, Bangladesh, pp. 261-266, December 2012.

[4] A.Vahdat and D. Becker. "Epidemic routing for partially connected ad hoc networks*". Technical Report CS-200006*, Duke University, USA, April 2000.

[5] K A. Lindgren, A. Doria, and O. Schelen, "Pobabilistic Routing in Intermittently Connected Networks," *SIGMOBILE Mobile Computing Communications Review*, vol. 7, no. 3, pp. 19-20, July 2003.

[6] J. Burgess, B. Gallagher, D. Jensen, and B. N. Levine, "Maxprop: Routing for Vehicle-Based Disruption-Tolerant Networks," *INFOCOM, IEEE*, pp. 1-11, April 2006.

[7] A. Balasubramanian, B. Levine, and A. Venkataramani, "Dtn Routing as a Resource Allocation Problem," *SIGCOMM Computing Communications Review.*, vol. 37, no. 44, pp. 373-384, August 2007.

[8] R. Ramanathan, R. Hansen, P. Basu, R. Rosales-Hain, and R. Krishnan, "Prioritized Epidemic Routing for Opportunistic Networks," *Proceedings of the 1st international MobiSys Workshop on Mobile Opportunistic Networking*, USA, pp. 62-66, June 2007.

[9] G. Sandulescu and S. Nadjm-Tehrani, "Opportunistic DTN Routing with Window-Aware Adaptive Replication," *Proceedings of the 4th Asian Conference on Internet Engineering*, USA, pp. 103-112, November 2008.

[10] T. Spyropoulos, K. Psounis, and C. S. Raghavendra, "Spray and Focus: Efficient Mobility-Assisted Routing for Heterogeneous and Correlated Mobility," *Proceedings of the Fifth IEEE International Conference on Pervasive Computing and Communications Workshops,* USA, pp. 79-85, March 2007.

[11] S. C. Nelson, M. Bakht, and R. Kravets, "Encounter-Based Routing in DTNs," *INFOCOM*, Rio de Janeiro, Brazil, pp. 846-854, April 2009.

[12] S.M.A. Iqbal, "Multishceme Spray and Wait Routing in Delay Tolerant Networks by Exploiting Nodes Delivery Predictability", *International Conference on Computer and Information Technology (ICCIT-2012)*, Chittagong, Bangladesh, pp. 255-260, December 2012.

[13] S.M.A. Iqbal, A.K. Chowdhury & A. Akhter, "AAR: Agent Aided Routing for Faster and Efficient Spreading of Messages in Delay Tolerant Networks", *Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON-2013)*, Krabi, Thailand, pp. 1-8, May 2013.

[14] J. Xue, X. Fan, Y. Cao, J. Fang, and J. Li, "Spray and Wait Routing Based on Average Delivery Probability in Delay Tolerant Network," *Proceedings of the International Conference on Network Security, Wireless Communications and Trusted Computing,* China, pp. 500-502, April 2009.

[15] V. Prueksasri, C. Intanagonwiwat, K. Rojviboonchai, "DNH-SaW: The Different Neighbor-History Spray and Wait Routing Scheme for Delay Tolerant Networks" *Proceedings of International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS, 2011),* Chiang Mai, Thailand, pp. 1-5, December 2011.

[16] Y. Kim, J. Koo, E. Jung, K. Nakano, M. Sengoku, and Y. Park, " Composite Methods for Improving Spray and Wait Routing Protocol in Delay Tolerant Networks" *International Symposium on Communications and Information Technologies (ISCIT, 2010),* Japan, pp. 1229-1234, March 2010.

[17] http://www.netlab.tkk.fi/tutkimus/dtn/theone/, One Simulator, available at HTTP: November 2010.

THE EFFECTS OF NON-INHIBITORY SERPIN MASPIN ON CELL MIGRATION USING AN ARTIFICIAL NEURAL NETWORK

M A Al-Mamun[1], MA Hossain[1], C Fall[1], MS Alam[2] and R Bass[3]
[1]Computational Intelligence Group, Faculty of Engineering and Environment,
University of Northumbria at Newcastle, UK
[2]Department of Applied Physics, Electronics and Communication Engineering,
University of Dhaka, Dhaka-1000, Bangladesh
[3]Department of Applied Sciences, Faculty of Health and Life Sciences,
University of Northumbria at Newcastle, UK

***Abstract*— Maspin (Mammary Serine Protease Inhibitor) is a non-inhibitory member of the serpin family of protease inhibitors that influences multiple cellular functions including adhesion, migration, and invasion in tumour malignancy. In this paper, we describe a computational model of the effects of exogenous maspin on cellular proliferation, migration and invasion. . To date, the whole cellular mechanisms of maspin impact on cellular behaviors have not been clearly defined in any computational model. In this paper, we have used a feedforward artificial neural network to define cell signaling and cellular automata approach to define the each and every cell behavior. Results show that maspin reduces migration and invasion by 10-40% and 15-30% respectively; confirmed by published *in vitro* data. This is the first attempt to model the effects of maspin using a computational model to verify *in vitro* data. This will provide new insights into the tumour suppressive properties of maspin and inform the development of novel cancer therapies.**

Keywords— maspin, computational model, artificial neural network, cell migration, tumour growth

I. INTRODUCTION

Tumour can be thought as a dynamic self-organizing complex bio-system, normally started from disruptions in genetic mutations in a single cell. In cases of metastasis, cell migration is an inevitable factor in tumour progression [1]. Cell migration results in the intra-vasation (entering the blood stream) and extravasation (exiting from the vessel to the new sites) of cells and leads to metastases. At the cellular level, these processes involve both intracellular and extracellular mechanisms to stimulate proliferation, cell-cell adhesion, cell migration, cell-extracellular matrix (ECM) interactions and angiogenesis. This paper reports a computational model which considers cell migration and proliferation to model their principal roles in tumour malignancy due to presence of maspin.

Maspin, alternatively called SERPINB5 in genome coding, is a member of the serine protease inhibitor superfamily [2]. Recent studies show that its expression is down-regulated in breast and prostate cancers, but over-expressed in pancreatic, gallbladder and colorectal cancers [3, 4]. Maspin decreases tumour growth and metastasis *in vivo* [2] and invasion *in vitro* [5, 6]. These effects are proposed in many *in vitro* and *in vivo* models. This diversity motivated us to build a computational model of the effects of maspin to show its potential engagement with multiple cellular phenomena using different computational techniques.

There are reports that maspin works both inside and outside of the cell [3, 4]. It is possible that extracellular maspin directly has influence on cell migration, adhesion and angiogenesis, while indirectly influencing tumour cell proliferation and apoptosis. Maspin has been reported to bind to the ECM [7]. Intracellular maspin binding partners have been identified, including the transcription factor IRF-6 [5, 8], and histone deacetylase 1 influencing the Bcl-2/Bax signal axis [9]. It has been shown that recombinant maspin reduces the invasive potential of MDA-MB-435 cells by 30-40% by altering their integrin profiles [10]. Recently, it has been suggested that the uPA/uPA receptor (uPAR) complex is involved in the localization and adhesive functions of maspin through a maspin-uPA-uPAR-β_1 integrin mega-complex, which regulates mammary epithelial cell adhesion [11]. It is shown in many studies that maspin localization in cancer cells is necessary for tumour suppression in different kinds of cancer [3, 4]. Maspin has been reported to bind to integrin cell adhesion receptors and maspin has a certain effects to β_1 integrin activation status [12]. Another novel finding is that an intact G-helix (mimicked by 15-mer peptide chemically) is required for the effect of mapin on cell migration [13]. In a recent literature, a computational model was developed which showed maspin effects computationally [14]. In this paper we have taken the migration raw data from a previous study [13], where it was reported that the G-helix is essential and sufficient for the influence of maspin on cell migration. This model will focus on the effects of exogenous maspin over the surrounding tumour cells. The main objective of this paper was to analyse proliferative and migratory effects of maspin computationally; to determine whether we could accurately build a model which included these cellular effects. The proposed model was implemented, tested and verified through a set of experiments to demonstrate the merits and capabilities of the scheme.

II. MATERIALS AND METHOD

A. The Model

Computational models serve as a model for oncologists and physicians, able to make predictions which can handle multiple dynamic interacting variables of any biological phenomena.

Last few decades, researchers have proposed many tumour growth models, but majority models considered all types of cancer cells as a same property. We considered the whole tumour mass as a single entity and defined the global parameters for every cell. There are some previous attempts to model tumour growth characteristics at the cellular level [15, 16, and 17]. In [16], cells were considered as individual entities; the decision making process for each cell based on oxygen, glucose concentration and hydrogen ion production using a neural network. In our previous model, we established an *in silico* model to calculate tumour mass with consideration of oxygen, glucose, extracellular matrix (ECM), cell-cell adhesion and cell movement as key micro environmental parameters. We also integrated information regarding protein expression, tumour growth constraints defined as growth promoters/inhibitors and a bioreductive drug (TPZ) transport model for hypoxic tumour cells [18]. Here we modified the neural network to include the effect of the presence/absence of maspin on tumour growth concentrating on proliferation and migration. There are some models which investigated cancer at tissue level [19, 20].

A tumour tissue was developed starting with one cell at the centre of a two dimensional grid. In this cellular automaton model, each grid element was either occupied by an abnormal cell or was empty. The grid elements had local concentrations of oxygen. Each tumour cell was treated as an individual entity and identified through its position coordinates $x(x_{x,}y_y)$. The cell was strongly influenced by the tumour microenvironment and had its own decision mechanism that determined its behaviour during the growth activity. The decision mechanism was implemented using an artificial neural network. The neural network was told the number of neighbours, the concentration of oxygen and presence/absence of maspin. The response of the network at its output nodes determined the decision or behaviour of the cell. The response was one of three cellular behaviours: proliferation, quiescence or apoptosis.

We considered normal tumour growth with an oxygen supply as previously reported in [15]. The concentrations of glucose and hydrogen ions did not vary in the *in vitro* experiments, so we have not considered hypoglycemia or acidity in building this model. The model considered the presence/absence of maspin, oxygen concentration was taken as constant as oxygen and other nutrients were constant throughout the *in vitro* experiment. The varying distance of cells from a blood vessel can cause heterogeneity in the tumour microenvironment and the tumour mass. The evolution of oxygen with respect to time was maintained by the following second order partial differential equations (PDEs) defined in [15, 16].

$$\frac{\partial O_2(x,t)}{\partial t} = D_{O_2}\Delta O_2(x,t) - f_{O_2}(x,t) \quad (1)$$

Where, D_{O_2} is the oxygen diffusion constant. For maspin, we formed a new diffusion equation for maspin in the model

$$\frac{\partial M(x,t)}{\partial t} = D_M\Delta M(x,t) - f_M(x,t) \quad (2)$$

Where, D_M is the maspin diffusion constant. The term $f_l(x,t)$ was the utilization or production function of oxygen and maspin *(l=O_2, M)* for each cell at a specific position *x* and at time *t* and is described in equation 2,

$$f_l(x,t) = \begin{cases} 0 \text{ no tumour cell at that grid point} \\ cr_l F(x) \text{ grid point occupied by a tumour cell} \end{cases} \quad (3)$$

Where, cr_l are the base consumption/production rates, *F(x)* is the modulated energy consumed by the cell located at the grid element *x* and calculated in equation 3. It was used to report the differences for the energy consumptions among different subclones.

$$F(x) = \max(k(R - T_r) + 1, 0.25) \quad (4)$$

Where, *k* is the strength of modulation, *R* is the response of the neural network and T_r is the target response. The term max (,0.25) shows that the cell's metabolism was at least a quarter of the base line consumption rate as considered by Anderson [15]. This function also ensures that the cell with the greatest network response will consume more nutrients.

The tumour cell density $n_0 = 0.0025^{-2} = 1.6 \times 10^5$ cells cm^{-2} because the cell resides on a 2D grid. All presented models discretized partial differential equations using standard five-point finite central difference formulas and used length scale $\Delta d = 0.0025$ (the rescaling of the length gives each square cell an area of $6.25 \times 10^{-6}\ cm^2$[16]) and time scale $\Delta d = 5 \times 10^{-4}$. The parameter values incorporated during the tumour development process are given in table 1.

TABLE I: PARAMETERS USED DURING TUMOUR GROWTH

Value	Definition	Value	References
D_{O2}	Oxygen diffusion constant	$1.8\times10^5 cm^2 s^{-1}$	[21]
D_M	Maspin diffusion constant	$1\times10^{-6} cm^2 s^{-1}$	[22]
n_o	Cancer cell density	1.6×10^5 cells cm^{-2}	[23]
cr_{O2}	Base oxygen consumption rate	2.3×10^{-16} mol $cells^{-1}\ s^{-1}$	[24]
k	Modulation strength	6	Model specific
T_r	Target response	0.675	Model specific

B. Neural Network

The neural network structure developed here for the tumour cell growth model was divided into three layers: an input layer *I*, a hidden layer *H* and an output layer *O*. The number of specific neighbours each cell had *n(x, t)* and its local chemical composition; oxygen $O_2(x,t)$ and $M(x,t)$ was fed as inputs into the network. The nodes in the different layers were connected with varying weights, determined by two matrices *w* and *W* and the nodes in the hidden and output layer are equipped with internal thresholds *θ* and *Φ*. Other details of the

neural network parameters were as before [16]. The value of the input layer was determined by the micro-environment of the cell, these values were then fed thought the network and produced a response in the output layer that determined the behaviour of the cell. The model checked the values at the first three output nodes and selected the greatest one as the life-cycle phenotype for the cell. The aim was that the response of the NN modelled the behaviour of real cancer cells, as defined previously [16]. The cell activated its movement when the movement node was a value greater than 0.5, a situation where available oxygen was reduced. If a cell underwent apoptosis, it was no longer updated and the grid point where the cell resided was considered empty in the next time step. For simplicity we did not include necrosis (death because of lack of nutrients) in the model. The weights and biases were chosen by an iterative process to get the desired neural network response.

Figure1 shows the basic structure of the feedforward neural network modeling the behavior of each tumour cell in the tumour microenvironment. Alternatively, NN is used as genotype to phenotype conversion. The input layer can be thought as cell surface receptors, weights as signaling strength of these receptors, hidden layer as regulatory genes and finally the output can be thought as behaviors of the cell. We used *in vitro* cell migration data for three kinds of cancer cell lines (MC7, DU145 and PC3). *In vitro* data was collected over 13 hours so our model was for the same time, during which the migration of two kinds of tumour cell proliferating and quiescent cells can be seen. The key objective was to build a generalized model to support the *in vitro* experiments with a quantitative mathematical analysis of the effects of maspin *in vitro*. Other model parameters details are as [16, 17, 18].

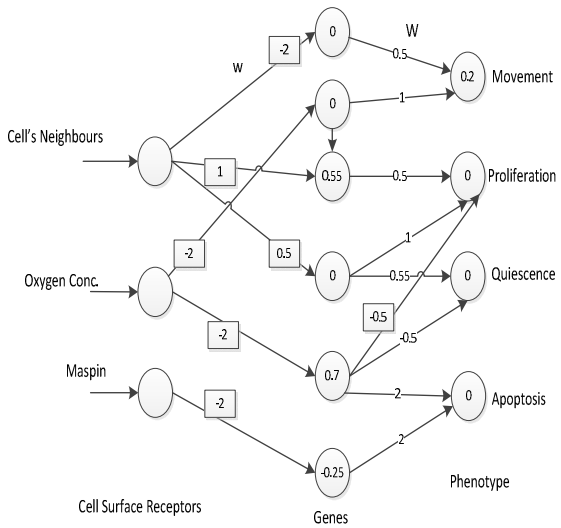


Fig 1. The basic structure of the neural network for tumour growth in the presence/absence of maspin

III. Results and Discussion

A. Initial Simulation

The two-dimensional tissue under consideration is represented by an $N \times N$ cellular automata grid. The initial simulation started from one cell at the centre of a grid, the cell divided and placed its daughter cell in one of the adjacent neighbourhoods after meeting some standards. At every new time step each cell performed one of its lifecycle actions (given in figure 2); proliferation, quiescence or apoptosis by selecting the highest neural network response; this involved the production/ consumption of oxygen/maspin. The cell consumed nutrients according to the selected phenotype; proliferating cells rapidly divide into daughter cells so they consumed more nutrients than the stationary ones. The daughter cells inherited attributes from their parents and generated their behaviour using these attributes as an input vector for the neural network. The number of mutations that occurred in the daughter cell network parameters was chosen from a Poisson distribution with parameter p. These mutations were then distributed equally over the network weights and nodes (see Fig. 1 and [16] for further details). The cell cycle decision mechanism was similar to previous models [16, 17]. The model was simulated with oxygen concentration as the determining factor of cell behaviour. In the model, the cell consumed maspin according to the selected phenotype of the neural network output (proliferation, quiescent, apoptosis and movement). Maspin was consumed via diffusion and the cell took maspin by the same process as for the normal tumour growth model. In order to make the simulations and analysis of the model simpler we non-dimensionalise Eqs (1)-(4) in the standard way (for details see [15]).

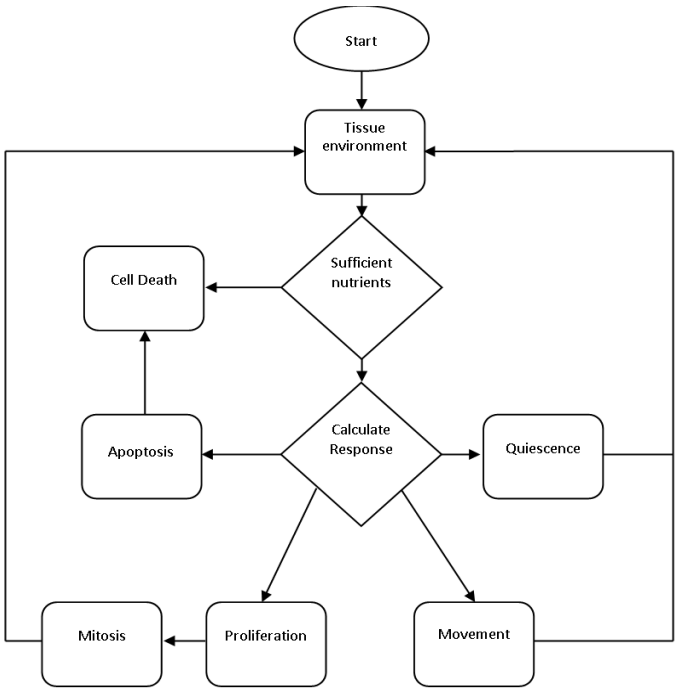


Fig 2. Life cycle actions of a tumour cell considered in the model

B. Cell Migration

The model allowed the migration of tumour cells with differential maspin expression during tumour growth to be visualized (Figure 3). The growth for different values of t up to 10 hours is presented. It can be seen that as time increased presence of maspin reduced the spread of the tumour cells as cell migration was reduced. This makes information about proliferation rates an important consideration. If the cells were proliferating at the same rate in presence and absence of

maspin, the tumour exposed to maspin would have an increased cell density. If there was an alteration in proliferation caused by the presence/absence of maspin this may not be the case.

In maspin growth model the tumour loses its circular shape after the second hour (t=20) of growth, indicating an effect of maspin in the model. After ten hours (t=100), this is enhanced making the maspin effects obvious when compared with the normal growth model. The tumour cells in the absence of maspin have formed a circular tumour and expanded to cover a greater surface area than those in the presence of maspin. In the case of the tumour cells exposed to constant maspin, we observed tumours grew with a compact round morphology similar to that reported for other models [16, 17] of proliferating and quiescent cells. Experimentally maspin reduced cell migration by up to 30 – 60% using different cancer cell lines *in vitro* [13]. We calculated the percentage of reduction in cell migration in our model as being up to 10-40%. Statistical two sample t- test has been used for analyzing the *in silico* data. There is good overlap between the measured and modeled migration data, the difference can be explained by considering maspin concentration. The different cell lines used for *in vitro* experiments expressed varying levels of maspin, including those that allowed an enhanced effect.

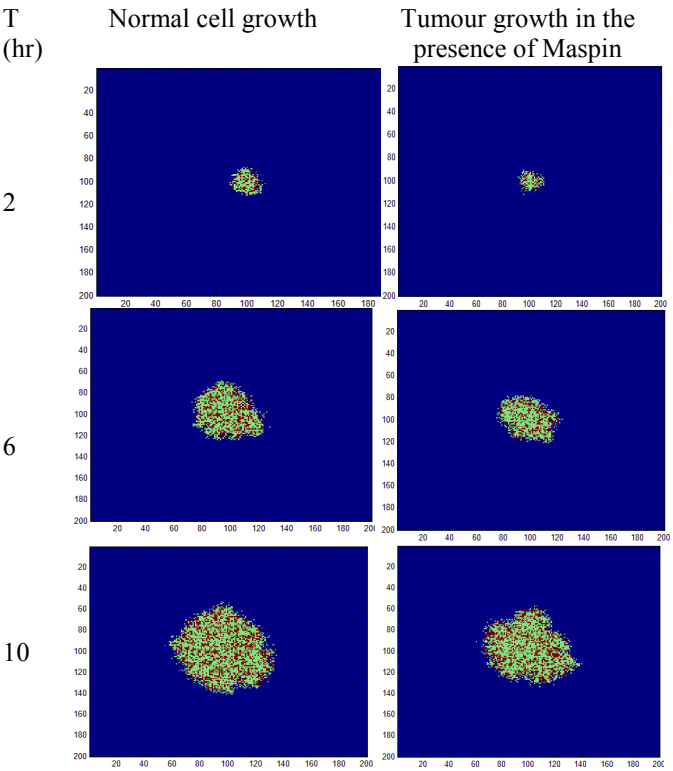


Fig 3. A comparison of cell migration and tumour growth in the presence and absence of maspin at t= 2,6,10 hours. Red cells indicate proliferating cells, yellow green is quiescent cells and blue space is empty area.

C. Invasion

The invasive distance achieved by the cells in both models was considered (Figure 4). Comparison between the two models showed that maspin reduced invasion by up to 30% which supports the *in vitro* hypothesis [13]. At each time point the cells in the presence of maspin had invaded less than those in the absence of it, this was enhanced after the 10th hour. The ability of tumours to invade is critical to their ability to metastasis and ultimately causes death. The loss of maspin as cancers progress has been suggested as the key event in development of metastatic prostate cancer [25]. This could be explained by the increase in invasive capacity of tumour cells in the absence of maspin as demonstrated here. The invasion curve presented here matched with the other literatures [14, 15].

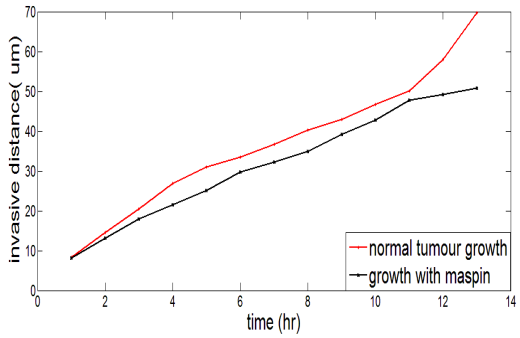


Fig 4. The modelled time evolution of invasive distance for both normal tumour growth and growth in the presence of maspin

D. Cell Proliferation

We considered the growth rate of the tumours in the presence of maspin in the models (Figure 5). The cell count demonstrated that maspin reduced cell proliferation in the model. This means that tumour density will not be increased in the presence of maspin, whereas overall the tumour will contain fewer cells. We vary the initial concentration of maspin (M_0) during the *in silico* experiments. It is shown from the results that when we increased the initial maspin concentration there produced less proliferating cells. We varied the concentration upto 10uM and results showed the consistent fashion. In figure 5, the results are shown for 0.5uM, 1.0 uM and 1.5uM initial maspin concentration. For three concentrations first two hours, the cell production was slower because of the long distance from the boundary. But from the third hour the curves were linearly increased. When tumour gets bigger, it reaches towards the boundary and they get more nutrients to grow quickly.

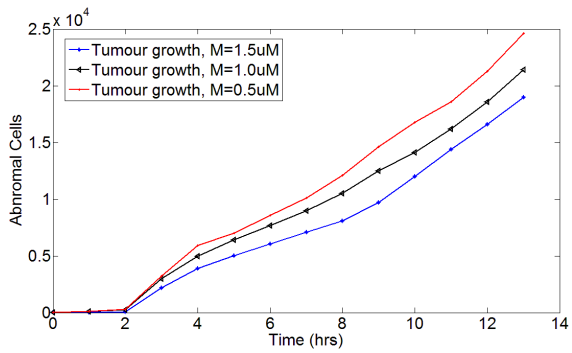


Fig 5.Tumour growth rate *in silico* experiments for initial maspin concentrations of 0.5uM, 1.0uM and 1.5uM

E. Comparative Analysis

For verifying the *in silico* experimental data we compared with *in vitro* data using different statistical techniques. We found that maspin reduced migration by 10-40% in comparison with normal growth. Comparison of the means between the control (*in vitro*) and experimental (*in silico*) stretch groups was determined by one-way ANOVA and Student's t test (Minitab16) with $p < 0.05$ considered significant.

Although this model does not take into account the whole range of effects of maspin, it can serve as a useful tool as it captures the tumour dynamics at the cellular level. This also presents the knowledge of the subcellular machinery focused on the relationship between environmental inputs and cellular phenotype. The effects of maspin were mimicked in this model using a decision neural network whose parameters served as the genetic material of the cell. The phenotype (outputs) were a function of the genotype (inputs), but also crucially depended on the microenvironment of the cell. Practically, it is still far from being a precise model of the subcellular dynamics, but this attempt brings the complete model a step closer. This model is basically extension of previous work with extended neural network with three inputs and also with *in vitro* data comparison [26]. This model can be compared with [14] and the results presented here extend the cell migration hypothesis in a simple manner. This model could be used to allow high throughput screening of variable conditions and physiological considerations as an alternative to biological experiments requiring expense, time and technical expertise.

IV. CONCLUSION

This paper has presented a cellular automaton model to show how proliferation and migration of tumour cells get affected by maspin. A computational model has been implemented, tested and verified through a set of experiments. Results showed that maspin reduces migration and invasion by 10-40% and 15-30% respectively. It is worth mentioning that the outcome of the proposed model resemble the *in vitro* experiments [13]. This proposed model took the biological hypothesis: maspin reduces cell migration for different types of cancer cell line and then mimicked the cell migration through computational modelling. The related literatures showed that it is possible to rebuild a computational model for maspin dynamics [14]. Our attempt was to show the simple cell migration reduction through the set of *in silico* experiments (fig 3). Our future work will be to develop a complete model to show the whole complex mechanism of the impact of maspin on cell-cell adhesion, cell-ECM interaction and apoptosis in subcellular level. Also, we will modify the neural network to demonstrate the efficiency with different hidden layers. Lastly, this kind of *in silico* model will enrich the possibility to establish maspin as a prognostic marker in near future.

REFERENCES

[1] Sahai, E., 2007. Illuminating the metastatic process. Nat. Rev. Cancer 7, 737–749.

[2] Zou Z, Anisowicz A, Hendrix MJ, Thor A, Neveu M, Sheng S, Rafidi K, Seftor E, Sager R., 1994. Maspin, a serpin with tumour-suppressing activity in human mammary epithelial cells. Science, 263, 526–529.

[3] Bodenstine, TM., Seftor, RE., Khalkhali-Ellis, Z., Seftor, EA., Pemberton, PA., Hendrix, MJ., 2012. Maspin: molecular mechanisms and therapeutic implications. Cancer Metastasis Rev. Dec; 31(3-4):529-51.

[4] Berardi R, Morgese F, Onofri A, Mazzanti P, Pistelli M, Ballatore Z, Savini A, De Lisa M, Caramanti M, Rinaldi S, Pagliaretta S, Santoni M, Pierantoni C, Cascinu S., Role of maspin in cancer. Clin Transl Med. 2013 Mar 7;2(1):8.

[5] Biliran, H., and Sheng, S., 2001. Pleiotrophic inhibition of pericellular urokinase-type plasminogen activator system by endogenous tumor suppressive maspin, Cancer Research, 61, 8676–8682.

[6] Sheng, S., Carey, J., Seftor, E. A., Dias, L., Hendrix, M. J. and Sager, R.,1996. Maspin acts at the cell membrane to inhibit invasion and motility of mammary and prostatic cancer cells. Proc Natl Acad Sci U S A, 93, 11669-74.

[7] Blacque, O.E. and Worrall, D.M., 2002. Evidence for a direct interaction between the tumour suppressor serpin, maspin, and types I and III collagen. J. Biol. Chem., 277, 10783–10788.

[8] Bailey, CM., Khalkhali-Ellis, Z., Kondo, S., Margaryan NVV., Seftor, REB., Wheaton, WW., Amir, S., Pins, MR., Schutte, BC., Hendrix, MJC., 2005. Mammary serine protease inhibitor (Maspin) binds directly to interferon regulatory factor 6: identification of a novel serpin partnership. J. Biol. Chem., 280, 34210-34210.

[9] Li, Z., Shi, HY., Zhang. HY., 2005. Targeted expression of maspin in tumour vasculatures induces endothelial cell apoptosis. Oncogene, 24, 2008-2019.

[10] Seftor, R.E., Seftor, E.A., Sheng, S., Pemberton, PA., Sager, R., and Hendrix, M.J., 1998. Maspin Suppresses the Invasive Phenotype of Human Breast Carcinoma", Cancer Res;58(24):5681-5.

[11] Endsley, M P., Yanqiu, Hu., Yong Deng et al., 2011. Maspin, the molecular bridge between the plasminogen activator system and β1 Integrin That Facilitates Cell Adhesion. J. Biol. Chem., vol. 286, NO. 28, pp. 24599–24607.

[12] Bass, R., Wagstaff, L., Ravenhill, L. & Ellis, V., 2009. Binding of extracellular maspin to beta1 integrins inhibits vascular smooth muscle cell migration. J. Biol. Chem., 284, 27712-20.

[13] Ravenhill, L., Wagstaff, L., Edwards, D.R., Ellis, V., and Bass, R., 2010. G-helix of Maspin mediates effects on cell migration and adhesion. J. Biol. Chem. Nov 19; 285(47):36285-92.

[14] Al-Mamun, M.A, Brown, L.J., M A Hossain, C Fall, L Wagstaff and Bass, R., 2013. A Hybrid Computational Model for the Effects of Maspin on Cancer Cell Dynamics, in Journal of Theoretical Biology, 337, pp 150–160.

[15] Gerlee, P. and Anderson, A.R.A., 2007.An evolutionary hybrid cellular automaton model of solid tumour growth. J. Theor. Biol. 246 (4), pp. 583-603.

[16] Kazmi, N., Hossain, MA., Phillips, RM., Al-Mamun, MA., Bass, R., 2012. Avascular tumour growth dynamics and the constraints of protein binding for drug transportation. J Theor Biol. Nov 21; 313:142-52.

[17] Kazmi, N., Hossain, MA., Phillips, RM., 2012. A hybrid cellular automaton model of solid tumor growth and bioreductive drug transport. IEEE/ACM Trans Comput Biol Bioinform. Nov-Dec;9(6):1595-606.

[18] Al-Mamun M.A., Kazmi, N., Hossain, M. A, Vickers, P and Jiang, Y., 2012. An intelligent decision support system for personalized cancer treatment. In: CIS2012: 11th IEEE Conference on Cybernetic Intelligent Systems, 23-24 August 2012, Limerick, Ireland.

[19] Alam, M.S., Hossain, M.A., Algoul, S., Majumader, M.A.A., Al-Mamun, M.A., Phillip,R. 2013. Multi-Objective Multi-drug Scheduling Schemes for Cell Cycle Specific Cancer Treatment, International Journal of Computers & Chemical Engineering, Vol-58, 11, p-14–32.

[20] Alam, N., Sultana, M., Alam, M. S., Al-Mamun, M.A.,Hossain, M.A., 2013. Optimal Intermittent Dose Schedules for Chemotherapy Using Genetic Algorithm, Advances in Distributed Computing and Artificial Intelligence Journal, Vol 1, No 5 .2013: SI#5, p-25-35.

[21] Grote, J., Susskind, S., and Vaupel, P., 1997. Oxygen diffusivity in tumour tissue (DS-carcinosarcoma) under temperature conditions within the range of 20-40 oC. Pflugers Architecture, 372, pp. 37-42.

[22] Young, M.E., Carroad, P.A., Bell, R.L., 1980. Estimation of diffusion coefficients of proteins. Biotechnol. Bioeng. 22(5) 947- 955.

[23] Casciari, J.J., Sotirchos, S.V. and Sutherland, R.M. 1988. Glucose diffusivity in multicellular tumour spheroids. Cancer Research, 48, pp. 3905–3909.

[24] Mueller-Klieser, W., Freyer, J.P., and Sutherland, R.M., 1986. Influence of glucose and oxygen supply conditions on the oxygenation of multicellular spheroids. British Journal of Cancer, 53 pp. 345-353.

[25] Luo, JL., Tan, W., Ricono, JM., Korchynskyi, O., Zhang, M., Gonias, SL., Cheresh, DA., and Karin, M., 2007. Nuclear cytokine-activated IKKalpha controls prostate cancer metastasis by repressing Maspin. Nature., 446, 690 – 4.

[26] Al-Mamun, M.A., Hossain, M.A., Alam, M.S., Bass, R., 2013. A Cellular automaton model of the effects of maspin on cell migration. Advances in Intelligent Systems and Computing Volume, 222, pp 53-60.

A Belief Rule Based (BRB) System to Assess Asthma Suspicion

Saifur Rahaman[1], Mohammad Shahadat Hossain[2]
[1,2]Computer Science and Engineering
University of Chittagong
Chittagong, Bangladesh
[1]saifurcubd@gmail.com, [2]hossain_ms@cu.ac.bd

***Abstract*—Asthma is a common chronic inflammatory disease. A belief rule based Clinical Decision Support System (CDSS) of asthma suspicion is enhancing the accuracy of suspicion. This research paper presents out the development of a Belief rule based (BRB) system to assess Asthma suspicion by using signs and symptoms. The recently developed generic belief rule-based inference methodology by using the evidential reasoning approach (RIMER) has been considered to develop this BRB system. This system can deal with various types of uncertainties, found in clinical sings, symptoms, and clinical domain knowledge. The knowledge base of this system has been constructed by taking account of real patient data and consultation with specialists. The practical case studies provided to test this system. It has been observed that the proposed model is effective and can generate better prediction than from manual system (usually carried out by specialist) in terms of accuracy.**

Keywords-Belief Rule Base (BRB); uncertainity; RIMER; Asthma; Suspicion; signs and symptoms

I. INTRODUCTION

Asthma is a common chronic disease with a high prevalence of approx and obstructive pulmonary diseases. The prevalence of asthma increases steadily over the latter part of the last century, first in the developed and then in the developing world[1].Now-a-days, many people of Bangladesh are suffering from Asthma. Current estimate suggest that asthma affects 300 million people world-wide and an additional 100 million people will be diagnosed by 2025[1]. Asthma is thought to be caused by a combination of genetic and environmental factors [2]. It also caused by a difficulty in breathing, normally associated with a severe lung condition. Anyone can get Asthma and it normally occurs with children with twice as many boys getting it than girls. The diagnosis of asthma is very complex and expensive. The suspicion of asthma is the preliminary process before diagnosis and treatment of Asthma. The suspicion process consists of observation of signs, and symptoms of a patient. However, this suspicion process contains errors since the signs, and symptoms. The reason for this is that various types of uncertainties such as vagueness, imprecision, incompleteness, ambiguity, ignorance are observed during the measurement of signs, and symptoms of a patient. For example, symptoms are expressed by a patient by using linguistic term and hence, it contains uncertainty due to vagueness, ambiguity and imprecision. Therefore, it is necessary to develop a computerized system to assess asthma suspicion by employing a methodology, which can handle the mentioned uncertainties. The cost-effectiveness of medical diagnosis and treatment is of growing importance [3] for the patient of Bangladesh. This RBR system would facilitate patients to avoid costly lab investigation.

In this paper, Section II presents an overview of Asthma along with signs, symptoms used to detect its suspicion and Asthma suspicion process. The review of related works is elaborated in Section III. Section IV introduces RIMER methodology. The prototype BRB system architecture, implementation strategy, knowledge base construction, and inference procedures are discussed in Section V. Section VI present the result and discussion, while Section VII concludes the paper.

II. ASTHMA

Asthma is a common chronic inflammatory disease of the airways characterized by variable and recurring symptoms, reversible airflow obstruction, and bronchospasm [4]. Airway inflammation contributes to airway hyper responsiveness, airflow limitation, respiratory symptoms, and disease chronicity. It is also characterized by an inflammation process, which induces bronchial hyper-responsiveness and usually reversible airway obstruction [5][6][7]. The signs and symptoms of Asthma in a patient characterize these. The main signs and symptoms discussed in next sub-section.

A. Main Signs and Symptoms of Asthma

The patients with Asthma may have a number of signs, and symptoms. The primary symptoms of Asthma are Cough and wheeze. In this research, five signs and symptoms have been considered. The most common signs and symptoms [1][2][4][8] are-

1) Cough 2) Breathlessness (Shortest of Breath)
3) Wheeze 4) Chest tightness
5) Respiratory Rate

The patient expresses these signs and symptoms by using linguistic terms and hence their expressions have various type of uncertainty like ambiguity, vagueness, and imprecision.

B. Asthma Suspicion Process

The diagnosis of Asthma consists assessment of suspicion and lab investigation or physical examination [4]. This paper deals with only suspicion stage of Asthma by using signs and symptoms. Manual Asthma suspicion system by using signs and symptoms of asthma with physician opinion. When a patient suspect to the Asthma then the patient is advised by the physician to go for lab investigation like pulmonary function

testing (SPIROMETRY)[1][9]. This paper present the development of a prototype CDSS to be used to assess posed to develop a prototype to assess asthma suspicion.

III. Review of Related Works

In the literature, different methodologies have been used to develop CDSS based expert system with uncertainties. The first well-known medical rule based expert system developed by Shortliffe [10][11] used to diagnosis bacterial infections. The rule based expert system, known as PERFEX [12] used to support to solve clinicians problems. INTERNIST-I[13] developed for diagnosis of complex problems in general internal medicine. However, these rule-based experts cannot handle uncertainties as mentioned in Section I. There are some CDSS expert systems such as AAPhelp-Leeds abdominal pain diagnosis system [14]. Environment and clinical work based CDDSs for Emergency Department, Intensive Care Unit and laboratory [15][16][17][18]. Some non-knowledge-based CDSSs which learn domain knowledge through large historical data [19]. Since uncertainties in both clinical domain knowledge and clinical data are inevitable, these CDDSs have not addressed this issue. However, Bayesian theory with a ranked list and posterior probability to each diagnosis was used in Iliad [20] and Fuzzy logic used in renal transplantation assignment [21]. The Bayesian theory and Fuzzy logic have their inherent limitations in representing all types of uncertainties. Artificial neural networks (ANNs) [22][23] are another popular methodology for modeling the CDSS. In ANN the relationships between inputs and output are not explicit or transparent, so it is difficult to understand or predict its behaviors. Therefore, the representation of Asthma suspicion with uncertain medical knowledge is crucial, requiring new methodologies and techniques [24]. Hence, this paper presents the employment of recently developed belief rule-base inference methodology using the evidential reasoning approach (RIMER) [25] to develop this prototype, which can handle various types of uncertainties from signs and symptoms in suspicion of Asthma.

IV. Overview of Rimer Methodology

The RIMER stands for Belief Rule-based Inference Methodology using the Evidential Reasoning approach [25]. This methodology consists of Belief Rule Base (BRB) (used to represent domain knowledge with uncertain information), the Evidential Reasoning (ER) [26] algorithm (used to deduce inference). This section presents the BRB structure and inference process with BRB using ER.

A. Domain Knowledge Representation using BRB

Belief Rules are the key constituents of a BRB, which include belief degree and are the extended form of traditional IF-THEN rules. In a belief rule, each antecedent attribute takes referential values and each possible consequent is associated with belief degrees [25]. The knowledge representation parameters are rule weights, antecedent attribute weights and belief degrees in consequents, which are not available in traditional IF-THEN rules. A belief rule can be defined in the following way.

$$R_k : \begin{cases} \text{IF } (P_1 \; is \; A_1^k) \wedge (P_2 \; is \; A_2^k) \wedge \dots\dots\dots \wedge P_{T_k} \; is \; A_{T_k}^k \\ \text{THEN } \{(C_1, \beta_{1k}), (C_2, \beta_{2k}), \dots\dots\dots, (C_N, \beta_{Nk})\} \end{cases} \quad (1)$$

$R_k : \left(\beta_{jk} \geq 0, \sum_{j=1}^{N} \beta_{jk} \leq 1 \right)$ with a rule weight θ_k , attribute weights $\delta_{k1}, \delta_{k2}, \delta_{k3}, \dots\dots, \delta_{kT_k}$ $k \in \{1, \dots\dots, L\}$

Where $\boldsymbol{P_1}, \boldsymbol{P_2}, \boldsymbol{P_3} \dots \boldsymbol{P_{Tk}}$ represent the antecedent attributes in the $\boldsymbol{k}$th rule. $A_i^k (i = 1, \dots\dots, T_k, k = 1, \dots\dots, L)$ represents one of the referential values of the $\boldsymbol{i}$th antecedent attribute $\boldsymbol{P_i}$ in the $\boldsymbol{k}$th rule. $\boldsymbol{C_j}$ is one of the consequent reference values of the belief rule. $\beta_{jk} (j = 1, \dots\dots, N, \; k = 1, \dots\dots, L)$ is the degree of belief to which the consequent reference value $\boldsymbol{C_j}$ is believed to be true. If $\sum_{j=1}^{N} \boldsymbol{\beta_{jk}} = \mathbf{1}$ the $\boldsymbol{k}$th rule said to be complete; otherwise, it is incomplete. $\boldsymbol{T_k}$ is the total number of antecedent attributes used in kth rule. $\boldsymbol{L}$ is the number of all belief rules in the rule base. $\boldsymbol{N}$ is the number of all possible consequent in the rule base.

An example of belief rule in Asthma suspicion BRB System can be written in the following way.

$$R_k : \begin{cases} \text{IF } (Cough \; is \; Yes) \wedge (Breathlessness \; is \; Moderate) \wedge \\ \quad (Wheeze \; is \; High) \wedge (Chest \; tightness \; is \; Yes) \wedge \\ \quad (Respiratory \; rate \; is \; Unrange) \\ \text{THEN } Asthma \\ \{(Severe, (0.80)), (Moderate, (0.20)), (Mild, (0.00)), (Normal, (0))\} \end{cases} \quad (2)$$

Where {(Severe, 0.80), (Moderate, 0.20), (Mild, 0.00), (Normal, 0.00)} is a belief distribution associated with Asthma consequent of the belief rule as represented in equation (2). The belief distribution states that the degree of belief associated with 'Severe' Asthma is 80%, 20% degree of belief associated with 'Moderate' Asthma, 0% degree of belief associated with 'Mild' Asthma while 0% degree of belief is associated with 'Normal' Asthma. Here, 'Severe', 'Moderate' 'Mild' and 'Normal' can be considered as the referential values of the consequent attribute "Asthma" of the belief rule. In this belief rule, the total degree of belief is (0.80+0.40+0.00+0.00)=1 and hence, the assessment is complete.

B. Inference System with BRB

The ER approach [21][26][27] developed to handle multiple attribute decision analysis (MADA) problem having both qualitative and quantitative attributes under uncertainty. Different from traditional MADA approaches, ER presents MADA problem by using a decision matrix, or a belief expression matrix, in which each attribute of an alternative described by a distribution assessment using a belief structure. The inference procedures in BRB inference system consists of various components such as input transformation, rule activation weight calculation, rule update mechanism, followed by the aggregation of the rules of a BRB by using ER.

The input transformation of a value of an antecedent attribute $\boldsymbol{P_i}$ consists of distributing the value into belief degrees of different referential values of that antecedent. This is equivalent to transforming an input into a distribution on referential values of an antecedent attribute by using their corresponding belief degrees. the The *ith* value of an antecedent attribute $\boldsymbol{P_i}(\boldsymbol{i} = \mathbf{1}, \dots\dots, \boldsymbol{T_k})$ at instant point in time can equivalently be transformed into a distribution over the

referential values, defined for the attribute by using their belief degrees [25]. The input value of $\boldsymbol{P_i}$, which is the **i**th antecedent attribute of a rule, along with its belief degree $\boldsymbol{\varepsilon_i}$ is shown below by equation (3). The belief degree $\boldsymbol{\varepsilon_i}$to the input value is assigned by the expert in this research.

$$H(P_i, \varepsilon_i) = \{(A_{ij}, \alpha_{ij}), j = 1, \cdots, j_i\},\ i = 1, \dots\dots, T_k \quad (3)$$

Here H is used to show the assessment of the belief degree assigned to the input value of the antecedent attribute. In the above equation $\boldsymbol{A_{ij}}$ (ith value) is the *j*th referential value of the input $\boldsymbol{P_i}$. . $\boldsymbol{\alpha_{ij}}$ is the belief degree to the referential value $\boldsymbol{A_{ij}}$ with $\alpha_{ij} \geq 0$. $\sum_{j=1}^{j_i} \alpha_{ij} \leq 1 (i = 1, \dots, T_k)$, and $\boldsymbol{j_i}$ is the number of the referential values. In this research, the input value of an antecedent attribute is collected from the patient or from the physician in terms of linguistic values. This linguistic value is assign degree of belief $\boldsymbol{\varepsilon_i}$ by taking account of expert judgment. This assigned degree of belief is then distributed in terms of belief degree $\boldsymbol{\alpha_{ij}}$ of the different referential values $\boldsymbol{A_{ij}}$ [Severe (H), Moderate (Mo), Mild, (M), Normal (N)] of the antecedent attribute.

It is important to note that in a rule (kth) only one of the referential values ($\boldsymbol{A_i^k}$, which is the element of $\boldsymbol{A_{ij}}$) of an antecedent attribute ($\boldsymbol{P_i}$) can be considered. Therefore, it is necessary to find out the degree of belief $\boldsymbol{\alpha_i^k}$ (which is the element of $\boldsymbol{\alpha_{ij}}$) of this referential value ($\boldsymbol{A_i^k}$), which can be defined as the matching degree at which the belief is matched. When matching degree is assigned to referential value of antecedent attributes then the rule is considered as activated. This phenomenon is called Packet Antecedent of a Rule. When the ***k***th rule is activated, the weight of activation of the ***k***th rule, $\boldsymbol{w_k}$ is calculated by using the flowing formula [25].

$$\boldsymbol{\omega_k} = \frac{\boldsymbol{\theta_k \alpha_k}}{\sum_{j=1}^{L} \boldsymbol{\theta_j \alpha_j}} = \frac{\boldsymbol{\theta_k} \prod_{i=1}^{T_k} (\alpha_i^k)^{\overline{\delta_{ki}}}}{\sum_{j=1}^{L} \boldsymbol{\theta_j} \left[\prod_{l=1}^{T_k} (\alpha_l^j)^{\overline{\delta_{jl}}}\right]} \text{ and } \overline{\boldsymbol{\delta_{ki}}} = \frac{\boldsymbol{\delta_{ki}}}{\boldsymbol{max_{i=1,\dots,T_k}\{\delta_{ki}\}}} \quad (4)$$

Where $\overline{\boldsymbol{\delta_{ki}}}$ is the relative weight of $\boldsymbol{P_i}$ used in the ***k***th rule, which is calculated by dividing weight of $\boldsymbol{P_i}$with maximum weight of all the antecedent attributes of the kth rule. By doing so, the value of $\overline{\boldsymbol{\delta_{ki}}}$ becomes normalize, meaning that the range of its value should be between 0 and 1. $\boldsymbol{\alpha_k} = \prod_{i=1}^{T_k} (\boldsymbol{\alpha_i^k})^{\overline{\delta_{ki}}}$ is the combined matching degree, which is calculated by using multiplicative aggregation function.

When the ***k***th rule as given in.(1) is activated, the incompleteness of the consequent of a rule can also result from its antecedents due to lack of data. An incomplete input for an attribute will lead to an incomplete output in each of the rules in which the attribute is used. The original belief degree $\overline{\beta_{ik}}$ in the ***i***th consequent $\boldsymbol{C_i}$ of the ***k***th rule is updated based on the actual input information as[15][25].

$$\boldsymbol{\beta_{ik}} = \overline{\boldsymbol{\beta_{ik}}} \frac{\sum_{t=1}^{T_k} \left(\tau(t,k) \sum_{j=1}^{J_t} \alpha_{tj}\right)}{\sum_{t=1}^{T_k} \tau(t,k)} \quad (5)$$

Where, $(t,k) = \begin{cases} 1, & if\ P_i\ is\ used\ in\ defining\ R_k (t = 1, \dots, T_k) \\ 0, & otherwise \end{cases}$

Here $\overline{\beta_{ik}}$ is the original belief degree and $\boldsymbol{\beta_{ik}}$ is the updated belief degree.

ER approach is used to aggregate all the packet antecedents of the $\boldsymbol{L}$ rules to obtain the degree of belief of each referential values of the consequent attribute by taking account of given input values $\boldsymbol{P_i}$ of antecedent attributes. This aggregation can be carried out either using recursive or analytical approach. In this research analytical approach has been considered since it is computationally efficient than recursive approach [21][26]. Using the analytical ER algorithm [27], the conclusion *O(Y),* consisting of referential values of the consequent attribute, is generated. Equation (6) as given below illustrates the above phenomenon. :

$$\boldsymbol{O(Y) = S(P_i) = \{(C_j, \beta_j),\ j = 1, \dots, N\}} \quad (6)$$

Where $\boldsymbol{\beta_j}$ denotes the belief degree associated with one of the consequent reference values such as $\boldsymbol{C_j}$. The $\boldsymbol{\beta_j}$ is calculating by analytical format of the ER algorithm [15][27] as illustrated in equation (7).

$$\boldsymbol{\beta_j} = \frac{\boldsymbol{\mu} \times \left[\prod_{k=1}^{L}\left(\left(\boldsymbol{\omega_k \beta_{jk} + 1 - \omega_k} \sum_{j=1}^{N} \boldsymbol{\beta_{jk}}\right)\right) - \prod_{k=1}^{L}\left(\boldsymbol{1 - \omega_k} \sum_{j=1}^{N} \boldsymbol{\beta_{jk}}\right)\right]}{\boldsymbol{1 - \mu} \times \left[\prod_{k=1}^{L} \boldsymbol{1 - \omega_k}\right]} \quad (7)$$

With

$$\boldsymbol{\mu} = \left[\sum_{j=1}^{N} \prod_{k=1}^{L} \left(\boldsymbol{\omega_k \beta_{jk} + 1 - \omega_k} \sum_{j=1}^{N} \boldsymbol{\beta_{jk}}\right) - (\boldsymbol{N} - \boldsymbol{1}) \times \prod_{k=1}^{n} \left(\boldsymbol{1 - \omega_k} \sum_{j=1}^{N} \boldsymbol{\beta_{jk}}\right)\right]^{-1}$$

The final combined result or output generated by ER is represented by $\{(C_1, \beta_1), (C_2, \beta_1), (C_3, \beta_1), \dots\dots\dots, (C_N, \beta_N)\}$, Where $\boldsymbol{\beta_j}$ is the final belief degree attached to the ***j***th referential value $\boldsymbol{C_j}$of the consequent attribute, obtained after combining all activated rules in the BRB by using ER. This output can be converted into crisp/numerical value by assigning utility score to each referential value of the consequent attribute [21][25].-

$$H(A^*) = \sum_{j=1}^{N} u(C_j)\, B_j \quad (8)$$

Where $H(A^*)$ is the expected score expressed as numerical value and $u(C_j)$ is the utility score of each referential value.

V. A BRB ASTHMA SUSPICION SYSTEM

The prototype BRB system architecture along with its components is presented in this section. Implementation strategy of the CDSS will also be introduced.

A. System Architecture and its Implementation Strategy

System architecture represents how its components consisting of input, process, and output are organized. It also considers the pattern of the system organization, known as architectural style. The BRB System developed for Asthma suspicion adopts the three-layer architectural style[15][28]. It contains interface layer (used to be get the input and produce system outputs), application layer (consisting of inference Procedures with BRB using RIMER) and data management layer (consisting of initial rule-base developed using BRB and facts including signs, symptoms of Asthma). Figure 1 illustrates the system architecture of the BRB System, which includes the components as discussed above..

It is important to note that in the prototype CDSS as illustrated in Figure 1, MySQL, which is a relational database has been used in back-end to store and manipulate initial BRB, which represents knowledge base of the system. MySQL as back-end relational database has been chosen since it is flexible and portable in any system environment.

Java application (J2SE), which is available in Netbean 7.2 has been used to develop the interface and the components

of the application layer. The reason for selecting is that Java application is platform independent, flexible and robust [29].

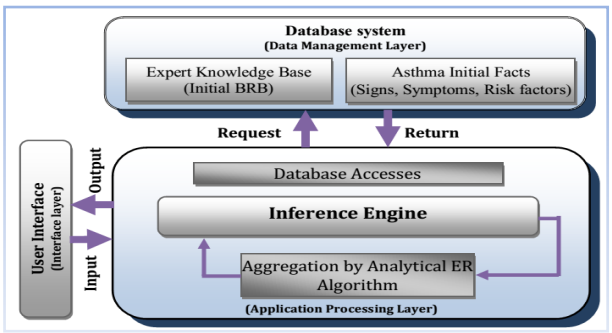


Figure 1. System architecture of the Asthma suspicion System

B System components

In this section the database, consisting of initial facts and BRB, and inference components of the prototype CDSS is elaborated.

a) Input Facts of Asthma Suspicion

The facts used to perform the Asthma suspicion are namely symptoms and signs as mentioned previously. Five symptoms and signs have been considered as input parameters for the BRB System. The input facts with clarification as illustrate in Table I. These input parameters have been considered as antecedent of the various rules used in the BRB System. The input values of these parameters are collected from either patient or physician and take an expert belief of corresponding input parameter.

TABLE I. THE INPUTS FACTS CLARIFICATION WITH EXPERT BELIEF

Sl. No.	Input Antecedent	Input	Expert belief Severity, (ε_i)
1	Cough	Yes	Severe , (1.0)
2	Cough	No	Normal, (0.2)
3	Breathlessness	No limitation	Normal, (0.0)
4	Breathlessness	Mild	Mild, (0.4)
5	Breathlessness	Moderate	Moderate , (0.7)
6	Breathlessness	Severe	Severe, (1.0)
7	Wheeze	High	Severe, (1.0)
8	Wheeze	Medium	Moderate , (0.6)
9	Wheeze	No	Normal, (0.0)
10	Chest tightness	Yes	Severe , (1.0)
11	Chest tightness	No	Mild, (0.3)
12	Respiratory rate	Range	Normal, (0.0)
13	Respiratory rate	Un range	Severe , (1.0)

For example, in Table 1, the input for parameter 'Chest tightness' has been obtained as "yes" and then a expert belief 1.0 and Severe has been taken by domain expert associated with this input parameter.

b) Knowledge base constructed using BRB

In order to construct the knowledge base for this BRB System, a BRB framework (by taking account of "British Guideline on the Management of Asthma" and "Guidelines for the Diagnosis and Management of Asthma: National Heart Lung and Blood Institute)[4][8] have been developed as illustrated in Figure 2.

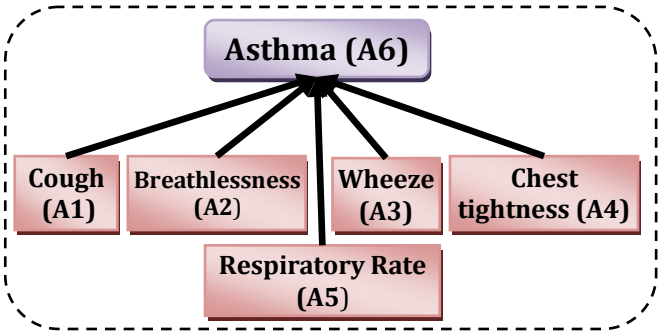


Figure 2. BRB Framework to suspect the Asthma

From the framework, it can be observed that input factors that determine suspicion include Cough (A1), Breathlessness (A2), Wheeze (A3), Chest tightness (A4), and Respiratory Rate (A5).

A BRB can be established in the following four ways[30]- 1) Extracting belief rules from expert knowledge 2) Extracting belief rules by examining historical data; 3) Using the previous rule bases if available, and 4) Random rules without any pre-knowledge. In this paper, we construct initial BRB by the domain expert knowledge. This BRB consists of only Asthma (A6) Rule base. It has five antecedent attributes. Each antecedent attribute consists of two, three, and four referential values. Hence, this entire BRB consists of 96 belief rules as illustrated in Table II. It is assumed that all belief rules have equal rule weight and all antecedent attributes have equal weight. The initial belief degree assigned to consequent of a rule has been carried out by taking opinion of two domain specialists. Four referential values such as Severe (S), Moderate (Mo), Mild (M) and Normal (N) have been considered for each consequent attribute of the belief rule.

An example belief rule of this BRB system is given below-

R1: ***IF*** (Cough is Yes) ***AND*** (Breathlessness is Severe) ***AND*** (*Wheeze* is High) ***AND*** (Chest tightness is Yes) **AND** (Respiratory rate is Un-range)

THEN Asthma is {S (1.00), Mo (0.00), M (0.00), N(0.0)}

TABLE II. INITIAL BELIEF RULE BASE FOR ASTHMA SUSPICION BRB SYSTEM

Rule ID	Rule Weight	IF Antecedent attributes (A1-A5)					Then Consequent Asthma (A6)			
		Cough (A1)	Breathlessness (A2)	Wheeze (A3)	Chest tightness (A4)	Respiratory Rate (A5)	Severe (S)	Moderate (Mo)	Mild (M)	Normal (N)
R1	1.00	Yes	Severe	High	Yes	Un range	1.000	0.000	0.000	0.000
R2	1.00	Yes	Severe	High	Yes	Range	0.600	0.000	0.000	0.400
R3	1.00	Yes	Severe	High	No	Un range	0.810	0.000	0.190	0.000
……	…..	…..	…….	…….	…..	……	……	……	……	…….
R95	1.00	No	No limitation	No	No	Un range	0.200	0.000	0.000	0.800
R96	1.00	No	No limitation	No	No	Range	0.000	0.000	0.000	1.000

In the above belief rule, the belief degrees attached to the three referential values of the consequent attribute. For example, Having 100 patients whose clinical information such as "Cough is Yes", "Breathlessness is Severe", "*Wheeze* is High", "Chest tightness is Yes", and "Respiratory rate is Un-range". It is meaning that clinical information of the antecedent parts of the rule is evaluated with high grade of the referential values of the corresponding antecedent attribute of the rule then all patients to be judged as 'Severe' to suspicion of Asthma. Consequently, it can assign the initial belief degrees to the consequents in the rule as "1" to "S", "0" to "Mo", "0" to "M", and "0" to "N" as shown in rule R1

c) Inference Engine

The backward and forward chaining inference procedures used in traditional CDSSs. The BRB System uses ER approach [25][27], as elaborated in section IV.B. The inference engine works in the following ways:

i) First reads input data, acquired from the patient or the physician
ii) The input data are then transformed
iii) Calculates activation weight of all rules using (4),
iv) Updates belief degree of consequence in rules using (5)
v) Finally, aggregates all rules by using (6)(7).

The result of Inference engine or system result in a numerical value generates by using (8), which will be discusses in next section. The inference engine has been developed by using J2SE in Netbean IDE.

VI. Result and Discussion

Patients with cough and shortness of breath (breathlessness) have been selected to assess suspicion of Asthma. The physician has collected the signs and symptoms of patient with cough and shortness of breath. The collected information of the patients have been used as input data in this prototype to assess suspicion of the Asthma as shown in Table III. Table III also illustrates the Asthma suspicion carried out by the physician and suspicion by this BRB System. The five patient's simulated input data and suspicion result have been shown in Table III.

It is important to note that the final consequent attribute "Asthma" of this BRB System, which measured the suspicion of Asthma have assigned four referential values namely "Severe", "Moderate" "Mild", and "Normal". Hence, in order to obtain crisp value or single numerical value in terms of percentage, specific utility score assigned to each referential value. For example, utility score 100% has been assigned to 'Severe', 66% assigned to 'Moderate', 33% assigned to 'Mild and 0% assigned to 'Normal'. The Asthma suspicion of this System measured in percentage for recommendation to patient, which generated using utility equation (8).

For example, an overall system output of Asthma suspicion is 91.23% can be obtained, by using degree of belief associated with referential values such as {(Severe, 0.81), (Moderate, 0.12), (Mild, 0.07), (Normal, 0)}.

If Asthma suspicion result is found more that 50% in this System then the result is considered as Asthma suspected positive (+) otherwise it is Asthma suspected negative (-) as illustrated in Table III. The five input parameters (as shown in Figure 2 and Table III) of this System collected from 40 patients having cough and shortness of breath to test this BRB System. The 40 patients have been considered for this research, although Table III shows the five patients for simplicity.

The Receiver Operating Characteristic (ROC) curve can be used to analyze effectively performances of Suspicion tests having ordinal or continuous results [31]. It can be used to test the results of BRB Expert System and manual system result by taking account of benchmark results (which is the actual Asthma recorded results of these patients). The performance of the system can be measured by calculating the Area Under Curve (AUC) [31][32][33][34].The large AUC is the more accurate and reliable results. Column 'Asthma Suspicion', 'Prototype Asthma Suspicion' and Manual System' of Table III shows the Actual Asthma record of the patient, The BRB System generated results of Asthma suspicion, and the manual system results of Asthma suspicion by taking account of same patient's data respectively. Figure 3 shows the two ROC curves, one represents the suspicion performances of the BRB System and another one is manual system.

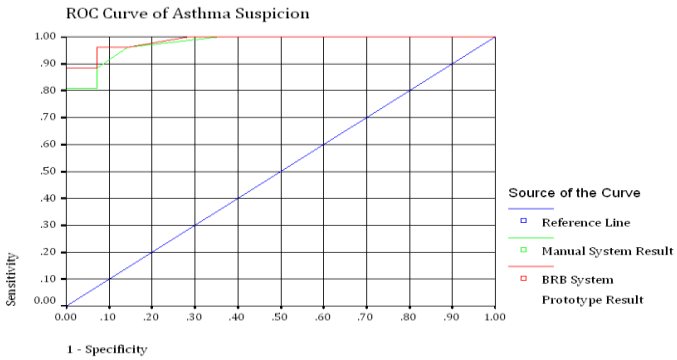


Figure 3. ROC curves illustrate the ACS suspicion performance of BRB Expert System and manual system

TABLE III. The Patient Information and Result of Asthma Suspicion of these patient

ID	Cough	Breathlessness	Wheeze	Chest tightness	Respiratory Rate	Asthma Suspicion	Prototype Asthma Suspicion	Manual System
P1	Yes	Moderate	High	Yes	Un range	+	93.2%	95%
P2	Yes	Severe	Medium	No	Range	-	73.2%	70%
P3	Yes	Mild	High	Yes	Un range	+	86.6%	80%
P4	Yes	Severe	No	Yes	Range	+	59.6%	65%
P5	Yes	No limitation	No	No	Range	-	30%	30%

The ROC curve with red line in Figure 3 illustrates the BRB system Asthma suspicion result while the curve with green line illustrates the manual system Asthma suspicion result. The AUC for BRB System is 0.986 (95% confidence intervals 0.960 – 1.012), and the AUC of manual system is 0.977 (95% confidence intervals 0.939 – 1.014). From the AUC of the two systems result, it can be observed that AUC of BRB System is greater than the AUC of manual system. This implies that

results generated by the BRB System is better than the results generated by the manual system, which uses traditional rule without taking account of uncertainty. The SPSS 11.5 has been used to construct the ROC curve and to calculate the AUC of these curves.

VII. Conclusion

Today‘s need to develop a strong expert system in clinical domain such as Asthma suspicion system, which combine clinical domain and IT to help reduce the medical error and various types of uncertainties and reduce medical cost. This paper presented the development and application of a BRB System for Asthma suspicion using signs and symptoms. This BRB System employed a novel methodology known as RIMER, allows the handling of various types of uncertainty found in clinical domain knowledge as well as in signs and symptoms. Hence, be considered as a robust tool can utilized in suspecting of Asthma. Most importantly, the system will play an important role in reducing the cost of lab investigations. The system will facilitate patients in taking precautionary measures well in advance. This Asthma suspicion BRB System can provide a percentage of suspicion recommendation, which is more reliable and informative than from the traditional physician’s opinion. This system performance is better than traditional (Manual) system. However, This BRB System has been developed only to assess suspicion Asthma from its signs and symptoms but cannot be used to diagnosis of Asthma. In future, an attempt will be taken to enhance the system with the capability to supporting the diagnosis of Asthma.

Acknowledgment

The authors are grateful to Dr. MD. Ismaeil, CMC, Chittagong, Bangladesh, for providing domain knowledge.

References

[1] Nicki R. Colledge, Brian R. Walker, Stuart H. Ralstonl, Davidson_s Principles and Practice of Medicine 21st Edition, Chapter 19, p662-667, ISBN-13: 978-0-7020-3084-0, 2010

[2] Martinez FD. "Genes, environments, development and asthma: a reappraisal". *Eur Respir J* **29** (1): 179–84, 2007

[3] Mark DB, Talley JD, Topol EJ, et al. Economic assessment of platelet glycoprotein IIb/IIIa inhibition for prevention of ischemic complica - tions of high-risk coronary angioplasty. EPIC Investigators. Circulation; 94:629–35, 1996

[4] Denise Dougherty, John Mitchell, Nancy Sander, Richard M. Nowak, Michael Schatz, William Busse, Guidelines for the Diagnosis and Management of Asthma, National Asthma Education and Prevention Program (NAEPP), US National Heart, Lung, and Blood Institute, 2007[A Asthma Guideline]

[5] Tattersfield AE, Knox AJ, Britton JR, Hall IP: Asthma. Lancet, 360:1313-1322, 2002

[6] Busse WW, Lemanske RF Jr, Asthma. N Engl J Med 344(5):350–362, 2001

[7] Goldman L, Ausiello D, Cecil textbook of medicine, 22nd ed. W. B. Saunders, Philadelphia, PA, 2004

[8] Graham Douglas, Dr Bernard Higgins, Neil Barnes, Anne Boyter, at al. British Guideline on the Management of Asthma (A national clinical guideline), ISBN 978 1 905813 28 5, January 2012

[9] Jones CA, Morphew T, Clement LT, Kimia T, Dyer M, Li M, et al. A school-based case identification process for Identifying inner city Child- ren with asthma: the Breathmobile program. Chest; 125 (3): 924-34, 2004

[10] Shortliffe, E. H. Computer-Based Medical Consultations: MYCIN, New York, Elsevier , 1976

[11] Buchanan and Shortliffe, Stanford University , MYCIN expert system, 1984

[12] Ezquerra et al., Emory University Hospital, 1992

[13] Kumar, K. A., Singh, Y. & Sanyal, S. (2009) Hybrid approach using case-based reasoning and rule-based reasoning for domain independent clinical decision support in ICU. Expert Systems with Applications,36,65-71

[14] De Dombal, F. T., Leaper, D. J., Staniland, J.R., Mccann, A. P. & Horrocks, J. C. Computer-aided diagnosis of acute abdominal pain. British Medical Journal,2(5804) 5 -9, 1972

[15] Rahaman, Saifur; Hossain, M.S. "A belief rule based clinical decision support system to assess suspicion of heart failure from signs, symptoms and risk factors", *Informatics, Electronics & Vision (ICIEV), 2013 International Conference,* page(s)1-6, DOI: 10.1109/ICIEV.2013.6572668

[16] Mack, E. H., Wheeler, D. S. & Embi, P. J. Clinical decision support systems in the pediatric intensive care unit. Pediatric Critical Care Medicine,10,23-28 , 2009

[17] Grams, R. R. Clinical laboratory test reference (CLTR). Journal of Medical Systems,17,59-67, 1993

[18] Kumar A, et al. Acute coronary syndromes: Diagnosis and management, part I. Mayo Clinic Proceedings. 2009; 84: 917.

[19] Berner, E. S. & La Lande, T. J., Overview of Clinical Decision Support Systems. IN BERNER, E. S. (Ed.) Clinical Decision Support Systems: Theory and Practice. 2nd ed. New York, Springer-Verlag, 2007

[20] Warner, H.R.J., 1989. Iliad: moving medical decision-making into new frontiers. Methods Inform. Med. 28, 370–372

[21] Yuan, Y., Feldhamer, S., Gafni, A., Fyfe, F., Ludwin, D., 2002. The development and evaluation of a fuzzy logic expert system for renal transplantation assignment: is this a useful tool? Euro. J. Oper. Res. 142, 152–173

[22] Zhang, J., & Chen, Y. Food sensory evaluation employing artificial neuralnetworks.Sensory Review, 17(2), 150–158, 1997.

[23] E- Bomio, M. Neural networks and the future of sensory evaluation. Food Technology, 52(8), 62–63, 1998

[24] Lin, L., Hu, P.J.-H., Liu Sheng, O.R., 2006. A decision support system for lower back pain diagnosis: uncertainty management and clinical evaluations. Decis. Support Syst. 42, 1152–1169

[25] Yang, J. B., Liu, J., Wang, J., Sii, H. S. & Wang, H. W. (2006) Belief rule-base inference methodology using the evidential reasoning approach -RIMER. IEEE Transactions on Systems Man and Cybernetics Part A-Systems and Humans, 36, 266-285

[26] Yang, J. B. & Sen, P. (1994) A general multi-level evaluation process for hybrid MADM with uncertainty. Systems, Man and Cybernetics, IEEE Transactions on,24,1458-1473

[27] Y.M. Wang, J.B. Yang, D.L. Xu, Environmental impact assessment using the evidential reasoning approach, European Journal of Operational Research 174 (2006) 1885 –1913

[28] Roger S. Pressman, Software engineering: a practitioner‘s approach.— 5th ed. p. cm.— (McGraw-Hill series in computer science) Includes index. ISBN 0 -07-365578-3, p373- 374

[29] Herbert schildt, JAVA: the complete reference, J2SE 5 edition, 2005, p10

[30] D.L. Xu, J. Liu, J.B. Yang, G.P. Liu, J. Wang, I. Jenkinson, J. Ren, Inference and learning methodology of belief-rule-based expert system for pipeline leak detection, Expert Systems with Applications 32 103–113, 2007

[31] Body, R. Clinical Decision Rules to Enable Exclusion of Acute Coronary Syndromes in the Emergency Department. Faculty of Health, Psychology and Social Care. Manchester, Manchester Metropolitan University, 2009

[32] Skalska, H. & Freylich, V. Web-bootstrap estimate of area under ROC curve. Austraian journal of statistics,35,325-330, 2006

[33] Metz, C. E. Basic principles of ROC analysis. Seminars in Nuclear Medicine,8,283-298, 1978

[34] Hanley, J. A. () The robustness of the "Binormal" assumptions used in fitting ROC curves. Medical decision making,8,197 -203, 1988

A New Hierarchical Interconnection Network for Future Generation Parallel Computer

Md. Rabiul Awal*, M.M. Hafizur Rahman*, and M. A. H. Akhand†

*Dept. of Computer Science, KICT, IIUM, 50728, Kuala Lumpur, Malaysia
†Dept. of Computer Science and Engineering, KUET, Khulna - 9203, Bangladesh
E-mail: rabiulawal1@gmail.com, hafizur@iium.edu.my, akhandkuet@yahoo.com

Abstract

*A **M**idimew connected **M**esh **N**etwork (**MMN**) is a **MI**nimal **DI**stance **ME**sh with **W**rap-around links network of multiple basic modules, in which the basic modules are 2D-mesh networks that are hierarchically interconnected for higher-level networks. In this paper, we present the architecture of the MMN, addressing of node, routing of message, and evaluate the static network performance of MMN, TESH, mesh, and torus networks. It is shown that the MMN possesses several attractive features, including constant degree, small diameter, low cost, small average distance, moderate bisection width, and high fault tolerant performance than that of other conventional and hierarchical interconnection networks. It is also shown that with the same node degree, arc connectivity, bisection width, and wiring complexity, the diameter and average distance of the MMN is lower than that of the TESH network.*

Keywords: Massively Parallel Computers, Interconnection Network, MMN, and Static Network Performance.

I. Introduction

The demand of computation power will never stop, it is increasing day by day. However, sequential computer steadily increases their speed to meet the computation demand, and it has already been reached saturated. Thus, the only way to meet the increasing demand of computation power to solve the grand challenge problems is to use parallel computers. Parallel computers provide computationally efficient techniques for the solution of large and difficult problem in a reasonable time [1]. To solve the grand challenge problems in many areas such as development of new materials and sources of energy, development of new medicines and improved health care, strategies for disaster prevention and mitigation, weather forecasting, and for scientific research including the origins of matter and the universe. This makes the current supercomputer changes into massively parallel computer (MPC) systems with thousands of node (Tianhe-2, Titan, Sequia, Kei, Cray XT5-HE [2]), that satisfy the insatiable demand of computing power. In near future, we will need computer systems capable of computing at the petaflops or exaflops level. To achieve this level of performance, we need MPC with tens of thousands or even millions of nodes [3]. It is predicted that the parallel systems of the next decade will contain 10 to 100 millions of nodes [4]. One crucial step on designing such a MPC systems is to determine the topology of the interconnection network, because the overall performance is significantly affected by the network topology [5–8]. Many recent experimental and commercial parallel computers use direct networks for low latency and high bandwidth of inter-node communication. For future MPC with millions of nodes, the large diameter of conventional topologies is completely infeasible. The hierarchical interconnection network (HIN) provides a cost-effective way in which several network topology can be integrated together [9]. Therefore, HIN is a plausible alternative way to interconnect the future MPC [9] systems. A variety of hypercube based HINs found in the literature, however, its huge number of physical links make it difficult to implement. To alleviate this problem, several k-ary n-cube based HIN have been proposed [10, 11]. However, the performance of these networks does not yields any obvious choice of an interconnection network for MPC. No one is clear winner in all aspect of network design.

Since reducing the diameter is likely to improve the performance of an interconnection network, the problem of designing interconnection network with low diameter with scalability of network size is still desirable [11]. A Tori connected mESH (TESH) network [12–14] is an HIN aiming for large-scale 3D MPC systems, consisting of multiple basic modules (BMs) which are 2D-mesh networks. The BMs are hierarchically interconnected by a 2D-torus to build higher level networks. The topology of the contemporary massively parallel computers use k-ary k-cube networks. No HIN draws the potential attention from the industry community yet. However, TESH network is a suitable network because of constant small node degree and scalability for modular expansion.

The main objective of this paper is to find a network which is suitable for interconnecting a large number of nodes while keeping small diameter. It has already been shown that a minimal distance mesh with wrap-around links (midimew) network is an optimal topology in the sense that there is no direct symmetric network of degree 4 with lower diameter or average distance [16]. To fulfill our objective, with this key motivation, we have replaced

the higher level 2D-torus of a TESH network by a 2D midimew network. To use the free ports in the periphery of the 2D-mesh network for higher level interconnection, we kept the basic module as 2D-mesh network as TESH network. The modified TESH network is a 2D-midimew networks of multiple basic modules (BMs), in which the BMs are 2D-mesh networks that are hierarchically interconnected for higher-level networks. Analogous to the TESH network, we called it **M**idimew-connected **M**esh **N**etwork (**MMN**). MMN is a hierarchical interconnection network, thus allowing exploitation of computation locality, as well as providing scalability up to a million of nodes. In this paper, we address the architectures of the MMN and evaluate its static network performance. The static network performance will be evaluated in terms of node degree, network diameter, cost, average distance, bisection width, and arc connectivity.

The remainder of the paper is organized as follows. In Section II, we briefly describe the basic architecture of the MMN. Addressing of nodes of the MMN and the routing of messages are discussed in Section III and Section IV, respectively. The static network performance of the MMN is discussed in Section V. Finally, in Section VI, we conclude this paper.

II. Interconnection of the MMN

The **M**idimew connected **M**esh **N**etwork (**MMN**) is a hierarchical interconnection network consisting of multiple basic modules (BM) that are hierarchically interconnected to form a higher level network. A $(2^m \times 2^m)$ BM consists of a 2D-mesh network of 2^{2m} processing elements (PE) having 2^m rows and 2^m columns, where m is a positive integer. Considering $m = 2$, a BM of size (4×4) is portrayed in Fig. 1. Each BM has 2^{m+2} free ports at the contours for higher level interconnection. All ports of the interior nodes are used for intra-BM connections. All free ports of the exterior nodes, either one or two, are used for inter-BM connections to form higher level networks. In this paper, BM refers to a Level-1 network.

Successive higher level networks are built by recursively interconnecting (2^{2m}) immediate lower level sub-networks in a $(2^m \times 2^m)$ **MI**nimal **DI**stance **ME**sh with **W**rap-around links *(midimew)* network. As depicted in Fig. 2, considering $(m = 2)$ a Level-2 MMN can be formed by interconnecting $2^{2\times 2} = 16$ BMs. Similarly, a Level-3 network can be formed by interconnecting 16 Level-2 sub-networks, and so on. Each BM is connected to its logically adjacent BMs. It is useful to note that for each higher level interconnection, a BM uses $4\times(2^q) = 2^{q+2}$ of its free links, $2(2^q)$ free links for diagonal interconnections and $2(2^q)$ free links for horizontal interconnections. Here, $q \in \{0, 1, ..., m\}$, is the inter-level connectivity. $q = 0$ leads to minimal inter-level connectivity, while $q = m$ leads to maximum inter-level connectivity. As shown in Figure 1, for example, the (4×4) BM has $2^{2+2} = 16$ free ports. If we chose $q = 0$, then $4(2^0) = 4$ of the free ports and their associated links are used for each higher level interconnection, 2 for horizontal and 2 for diagonal interconnection. Among these 2 links, one is used for incoming link and another one for used for outgoing link, i.e., a single links is used for diagonal_in, diagonal_out, horizontal_in, and horizontal_out.

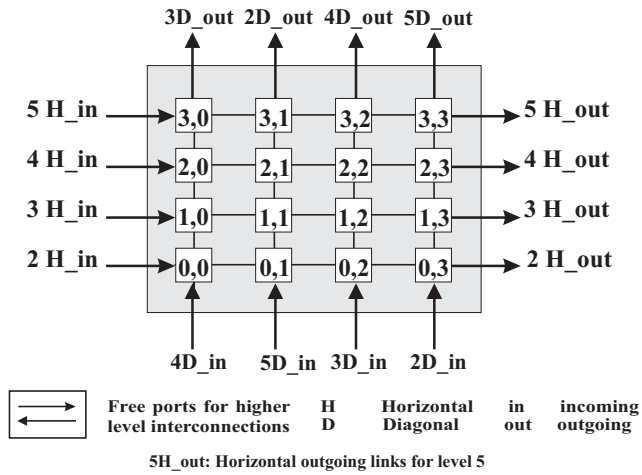


Fig. 1. Basic module of the MMN

A $\mathrm{MMN}(m, L, q)$ is constructed using $2^m \times 2^m$ BMs, has L levels of hierarchy with inter-level connectivity q. In principle, m could be any positive integer value. However, if $m = 1$, then the network degenerates to a hypercube network. Hypercube is not a suitable network, because its node degree increases along with the increase of network size. If $m = 2$, then it is considered the most interesting case, because it has better granularity than the large BMs. If $m \geq 3$, the granularity of the family of networks is coarse. If $m = 3$, then the size of the BM becomes (8×8) with 64 nodes. Correspondingly, the Level-2 network would have 64 BMs. In this case, the total number of nodes in a Level-2 network is $N = 2^{2\times 3\times 2} = 4096$ nodes, and Level-3 network would have 262144 nodes. Clearly, the granularity of the family of networks is rather coarse. In the rest of this paper we consider $m = 2$, therefore, we focus on a class of $\mathrm{MMN}(2, L, q)$ networks.

The highest level network which can be built from $(2^m \times 2^m)$ BM is $L_{max} = 2^{m-q} + 1$. With $q = 0$ and $m = 2$, $L_{max} = 5$, Level-5 is the highest possible level. The total number of nodes in a MMN having $(2^m \times 2^m)$ BMs is $N = 2^{2mL}$. Using maximum level of hierarchy, $L_{max} = (2^{m-q} + 1)$, the maximum number of nodes which can be interconnected by a $\mathrm{MMN}(m, L, q)$ is $N = 2^{2m(2^{m-q}+1)}$. For the case of (4×4) BM with $q = 0$, a network consists of 1 million nodes.

The question may arise, whether we need massively parallel computers with thousands of nodes or millions of nodes. The answer is 'yes'. Solving the most challenging problems in many areas of science and engineering, such as

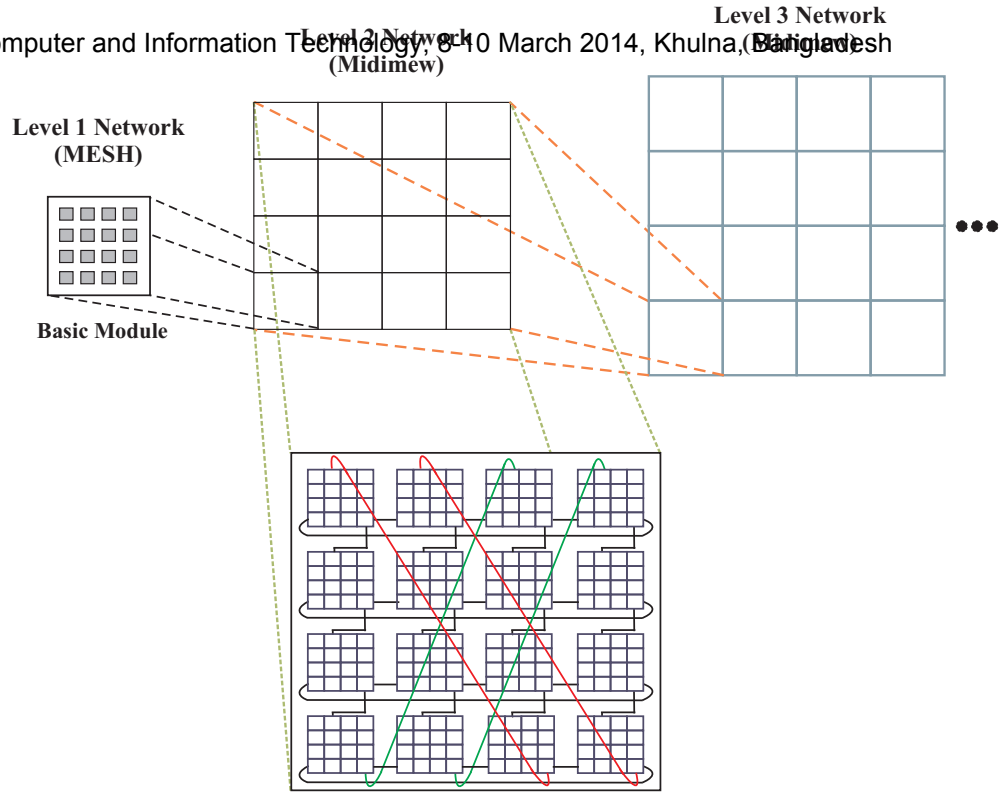


Fig. 2. Higher level network of a MMN

defense (maintaining national security), aerospace (space exploration and shuttle operation), disaster management (recovering from natural disaster), and weather forecasting (predicting and tracking severe weather), requires teraflop performance for more than a thousand hours at a time. This is why, in the near future, we will need computer systems capable of computing at the tens of petaflops level or even exaflops level. To achieve this level of performance, we need MPC system with thousands or millions of nodes.

III. Addressing of Nodes

Base-4 numbers are used for convenience of address representation. As seen in Figure 1, nodes in the BM are addressed by two digits, the first representing the row index and the next representing the column index. More generally, in a Level-L MMN, the node address is represented by:

$$\begin{aligned} A &= A^L A^{L-1} A^{L-2} \ldots \ldots \ldots A^2 A^1 \\ &= a_{n-1} a_{n-2} a_{n-3} \ldots \ldots \ldots a_2 a_1 a_0 \\ &= a_{2L-1} a_{2L-2}\ a_{2L-3} a_{2L-4} \ldots \ldots \ldots a_3 a_2\ a_1 a_0 \\ &= (a_{2L-1}\ a_{2L-2})\ (a_{2L-3}\ a_{2L-4})\ \ldots \ldots \\ &\quad \ldots \ldots \ldots (a_3\ a_2)\,(a_1\ a_0) \end{aligned} \tag{1}$$

Here, the total number of digits is $n = 2L$, where L is the level number. A^L is the address of level L and $(a_{2L-1} a_{2L-2})$ is the co-ordinate position of Level-$(L-1)$ for Level-L network. Pairs of digits run from group number 1 for Level-1, i.e., the BM, to group number L for the L-th level. Specifically, l-th group $(a_{2l-1}\ a_{2l-2})$ indicates the location of a Level-$(l-1)$ sub-network within the l-th group to which the node belongs; $1 \le l \le L$. In a two-level network the address becomes $A = (a_4\ a_3)\ (a_1\ a_0)$. The first pair of digits $(a_4\ a_3)$ identifies the BM to which the node belongs, and the last pair of digits $(a_1\ a_0)$ identifies the node within that BM.

The assignment of inter-level ports for the higher level networks has been done quite carefully so as to minimize the higher level traffic through the BM. The address of a node n^1 encompasses in BM_1 is represented as $n^1 = \left(a^1_{2L-1} a^1_{2L-2} \ldots \ldots \ldots a^1_3 a^1_2\ a^1_1 a^1_0\right)$. The address of a node n^2 encompasses in BM_2 is represented as $n^2 = \left(a^2_{2L-1} a^2_{2L-2} \ldots \ldots \ldots a^2_3 a^2_2\ a^2_1 a^2_0\right)$. The node n^1 in BM_1 and n^2 in BM_2 are connected by a link if the following condition is satisfied.

$\exists i\{a^1_i = (a^2_i \pm 1) \bmod 2^m\ \wedge \forall j(j \neq i\ \rightarrow\ a^1_j = a^2_j)\}$
where $i\%2 = 0,\ i, j \ge 2$
$\exists i\{a^1_i = (a^2_i \pm 1)\ \wedge \forall j(j \neq i\ \rightarrow\ a^1_j = a^2_j)\}$
where $a^1_i = (2^m - 1),\ i\%2 = 1,\ i, j \ge 2$
$\exists i\{a^1_i = (a^2_i \pm 1) \bmod 2^m \wedge \forall j(j \neq i \rightarrow a^1_j = a^2_j + 2)\}$
where $i\%2 = 1,\ i, j \ge 2$

IV. Routing Algorithm for MMN

Routing of messages in the MMN is performed from top to bottom as in TESH network [14, 15]. That is, it is first done at the highest level network; then, after the packet reaches its highest level sub-destination, routing continues within the sub-network to the next lower level sub-destination. This process is repeated until the packet arrives at its final destination. When a packet is generated at a source node, the node checks its destination. If the packet's destination is the current BM, the routing is performed within the BM only. If the packet is addressed to another BM, the source node sends the packet to the outlet node which connects the BM to the level at which the routing is performed. At each level, horizontal routing is performed first when the packet reaches the destined column, then diagonal routing start to reach the packet in the destination.

Let us consider an example in which a packet is to be routed from source node 000000 to destination node 130303. In this case, routing is to be done at Level-3, therefore the source node sends the packet to the outlet node of Level-3, 00 30 30, whereupon the packet is routed at Level-3, as shown in Figure 3. To avoid clutter diagonal wrap-around links are not shown in the figure. After the packet reaches the Level-2_$(1,1)$ network, then routing within that network continues until the packet reaches the $\mathrm{BM}(1,0)$. Finally, the packet is routed to its destination.

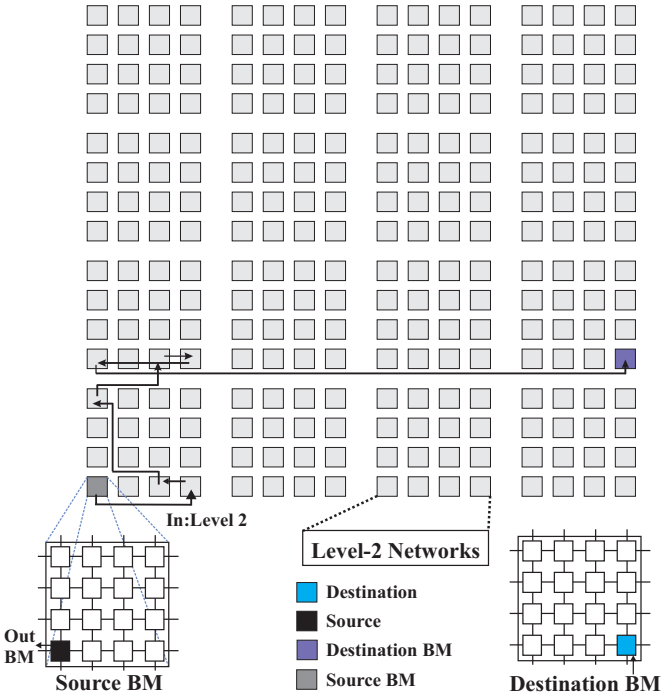


Fig. 3. An example of routing algorithm of the MMN

V. Static Network Performance

Although the actual performance of a network depends on many technological and implementation issues, several topological properties and performance metrics can be used to evaluate and compare different network topologies in a technology-independent manner. Most of these properties are derived from the graph theoretic model of the topology. In this section, we discuss some graph theory metrics that characterize the cost and performance of an interconnection network. These parameters have a direct impact on network performance. For the performance evaluation, we have considered mesh, torus, TESH network, and proposed MMN.

A. Node Degree

The *node degree* is defined as the maximum number of physical links emanating from a node. Since each node of the network has four links, the degree of the MMN is 4, and it is independent of network size. Constant degree networks are easy to expand, and the network interface cost of a node remains unchanged with increasing network size. The I/O interface cost of a particular node is proportional to its degree. It is shown in Table I that the degree of the mesh, torus, TESH, and MMN are equal, it is 4.

B. Diameter

The *diameter* of a network is the maximum inter-node distance, i.e., the maximum number of links that must be traversed to send a message to any node along the shortest path. As a definition, the distance between adjacent nodes is unity. Diameter is the maximum distance among all distinct pairs of nodes along the shortest path. The diameter is commonly used to describe and compare the static network performance of the interconnection network topology. Networks with small diameters are preferable. The smaller the diameter of a network, the shorter the time to send a message from one node to the node farthest away from it. Also, network diameter determines the largest base latency.

We have evaluated the diameter of the MMN and TESH network by simulation and mesh and torus network by their static formula and the result is tabulated in Table I. Clearly, the MMN has a much smaller diameter than that of TESH and mesh networks; and a slightly high diameter than that of torus network.

C. Cost

Inter-node distance, message traffic density, and fault-tolerance are dependent on the diameter and the node degree. The product $(diameter \times node\ degree)$ is a good criterion for measuring the relationship between cost and performance of a parallel computer systems [17]. An interconnection network with a large diameter has a very low message passing bandwidth, and a network with a high node degree is very expensive. In addition, a network should be easily scalable; there should be no changes in the basic node configuration as we increase the number of nodes. The cost of different networks is plotted in Table

Table I. Comparison of static network performance of various networks

	Node Degree	Diameter	Cost	Average Distance	Arc Connectivity	Bisection Width	Wiring Complexity
2D-Mesh	4	30	120	10.67	2	16	480
2D-Torus	4	16	64	8	4	32	512
TESH(2,2,0)	4	21	84	10.47	2	8	416
MMN (2,2,0)	4	17	68	9.07	2	8	416

I, and it is shown that the cost of MMN is far lower than that of mesh and TESH networks, and slightly higher than that of torus network.

D. Average Distance

The *average distance* is the mean distance between all distinct pairs of nodes in a network. Average distance is reflected on the average network latency. A small average distance results small communication latency, especially for distance-sensitive routing, such as store and forward. But it is also crucial for distance-insensitive routing, such as wormhole routing, since short distances imply the use of fewer links and buffers, and therefore less communication contention. We have evaluated the average distances for MMN and TESH network by simulation and mesh and torus networks by their corresponding formulas and the result is tabulated in Table I. It is shown that the average distance of MMN is lower than that of mesh and TESH networks, and higher than that of torus network.

Although the dynamic communication performance of a program on a parallel computer depends on the actual times taken for data transfer, a smaller average distance and diameter of an interconnection network yields a smaller communication latency of that network.

E. Bisection Width

The *Bisection Width (BW)* of a network is defined as the minimum number of links that must be removed to partition the network into two equal halves. Many problems can be solved in parallel using *binary divide-and-conquer:* split the input data set into two halves, and solve them recursively on both halves of the interconnection network in parallel, then merge the results from both halves into the final result. Small bisection width implies low bandwidth between the two halves, and it can slow down the final merging phase. On the other hand, a large bisection width is undesirable for the VLSI design of the interconnection network, since it implies a lot of *extra chip wires*, such as in hypercube.

Bisection bandwidth reflects the communication capacity between any two halves of the network when dealing with uniform, non-uniform, local, and non-local traffic. It is calculated by counting the number of links that need to be removed to partition the Level-L MMN. We have calculated the average distance of MMN, TESH, mesh, and torus networks by their respective static formula and it is shown in Table I. It is shown that the the bisection width of the MMN is exactly equal to that of the TESH network.

F. Arc Connectivity

Arc Connectivity measures the robustness of a network. It is a measure of the multiplicity of paths between nodes. Arc connectivity is the minimum number of links that must be removed in order to break the network into two disjoint parts. High arc connectivity improves performance during normal operation by avoiding link congestion, and also improves fault tolerance. The ratio between arc connectivity and the degree of a node gives a measure of static fault tolerance performance. A network is maximally fault-tolerant if its connectivity is equal to the degree of the network. The arc connectivity of various networks is shown in Table I. Clearly, the arc connectivity of the MMN is exactly equal to that of TESH network. However, the arc connectivity of the torus network is exactly equal to its degree. Thus, torus is more fault tolerant than all the networks. MMN is exactly equal fault tolerant to that of mesh and TESH networks.

G. Wiring Complexity

The *wiring complexity* of an interconnection network refers to the total number of links required to form the network. The wiring complexity depends on the node degree and it has a direct correlation to hardware cost and complexity. A (16×16) 2D-mesh and 2D-torus networks have $\{N_x \times (N_y - 1) + (N_x - 1) \times N_y\} = 16 \times (16-1) + (16-1) \times 16 = 480$ and $(2 \times N_x \times N_y = 2 \times 16 \times 16) = 512$ links, respectively. N_i represents the number of switches in the ith dimension. The wiring complexity of a Level-L TESH and MMN networks is $\left[\# \text{ of links in a BM} \times k^{2(L-1)} + \sum_{x=2}^{L} 2(2^q) \times k^{2(L-1)}\right]$. Considering, $m = 2$, a BM of MMN and TESH network have 24 links. Hence the total number of links of a Level-2 MMN and TESH are 416. Table I compares the wiring complexity of a MMN with that of several other networks. It is shown that the total number of links of MMN is lower than that of mesh and torus network and exactly equal to that of TESH network.

From the Table I, The question may arise, whether we need MMN over torus, where torus network results lower diameter and average distance to that of MMN. The answer is 'yes'. The torus network has a large number of long

wrap-around links. The operating speed of a network is limited by the physical length of its links. The length of the longest wire may become more important than the diameter of the network. The only drawback of MMN is that it has 4 long wire which are diagonally connected. Whereas for a 16×16 torus network, there are 32 long wrap-around links; among them 16 for x-direction and 16 for y-direction.

VI. Conclusion

A new hierarchical interconnection network, called **M**idimew connected **M**esh **N**etwork (**MMN**), is proposed for the high performance MPC systems. The architecture of the MMN, addressing of nodes, and routing of message were discussed in detail. We have evaluated the static network performance of the MMN, as well as that of several other interconnection networks. From the static network performance, it has been shown that the MMN possesses several attractive features, including constant node degree, small diameter, low cost, small average distance, and better bisection width. We have seen that with the same node degree, arc connectivity, bisection width, and wiring complexity, the diameter and average distance of the MMN is lower than that of the TESH network. Also, the diameter and average distance of the MMN is lower than that of mesh network and slightly higher to that of MMN. The MMN yields better static network performance with reasonable cost, which are indispensable for future generation high-performance MPC.

This paper focused on the architectural structure and static network performance. Issues for future work include the following: (1) evaluation of static network performance considering different value of m, L, & q and (2) evaluation of dynamic communication performance using dimension order routing algorithm.

Acknowledgment

This work is partly supported by FRGS13-065-0306, Ministry of Education, Government of Malaysia. The authors would like to thank the anonymous reviewers for their constructive comments and suggestions on the paper which have helped to improve the quality of the paper.

References

[1] B.S.P. Mishra and S. Dehuri, Parallel Computing Environments: A Review, *IETE Technical Review*, vol. 28, no. 3, pp. 240-247, 2011.

[2] http://top500.org/blog/lists/2013/06/press-release/

[3] V. Puente, J.A. Gregorio, R. Beivide and F. Vallejo, A Low Cost Fault Tolerant Packet Routing for Parallel Computers, *Proceedings of the 17th IEEE/ACM International Parallel and Distributed Processing Symposium (IPDPS)*, April, 2000.

[4] P. Beckman, Looking Toward Exascale Computing, *Keynote Spech. International Conference on Parallel and Distributed Computing, Applications and Technologies (PDCAT08)*, Dunedin, New Zealand, December 2008.

[5] Y. Yang, A. Funahashi, Akiya Jouraku, H. Nishi, H. Amano, T. Sueyoshi, Recursive Diagonal Torus: An Interconnection Network for Massively Parallel Computers, *IEEE Transactions on Parallel and Distributed Systems*, vol. 12, no. 7, pp. 701715, 2001.

[6] W.J. Dally, "Performance Analysis of k-ary n-cube Interconnection Networks," *IEEE Trans. on Computers*, vol. 39, no. 6, pp. 775-785, June 1990.

[7] M.M. Hafizur Rahman and S Horiguchi, Network Performance of Hierarchical Torus Network *Proc. of the 3rd PDCAT*, pp. 122–129, Kanazawa, Japan, 2002.

[8] M.M. Hafizur Rahman, X Jiang, MSA Masud, S Horiguchi, Network Performance of Pruned Hierarchical Torus Network *Proc. of the 6th IFIP Int'l Conference NPC*, pp. 9–15, Australia, 2009.

[9] M. Abd-El-Barr and T.F. Al-Somani, Topological Properties of Hierarchical Interconnection Networks: A Review and Comparison, *Journal of Electrical and Computer Engineering*, Hindawi Publishing Corporation, Vol. 2011, 12 pages.

[10] P.L. Lai, H.C. Hsu, C.H. Tsai, I.A. Stewart, A class of hierarchical graphs as topologies for interconnection networks, *Theoretical Computer Science, Elsevier*, Vo. 411, pp. 2912–2924, 2010.

[11] Youyao Liu, Cuijin Li, and Jungang Han, RTTM: A New Hierarchical Interconnection Network for Massively Parallel Computing, *Proc. of the HPCA*, LNCS 5938, pp. 264–271, 2010.

[12] V.K. Jain, T. Ghirmai, and S. Horiguchi, "TESH: A new hierarchical interconnection network for massively parallel computing", *IEICE Trans. on Inf. & Syst.*, vol.E80-D, no.9, pp.837-846, 1997.

[13] V.K. Jain and S. Horiguchi, "VLSI Considerations for TESH: A New Hierarchical Interconnection Network for 3-D Integration", *IEEE Trans on VLSI Systems*, vol.6, no. 3, pp. 346-353, 1998.

[14] M.M. Hafizur Rahman, Yasushi Inoguchi, Yukinori Sato, Yasuyuki Miura, and Susumu Horiguchi, "Dynamic Communication Performance of the TESH Network under Nonuniform Traffic Patterns", *Journal of Networks*, vol.4, no. 10, pp. 941 - 951, 2009.

[15] Faiz Al Faisal and M.M. Hafizur Rahman, "Symmetric Tori Connected Torus Network", *Proceedings of 2009 12th International Conference on Computer and Information Technology (ICCIT 2009)*, pp. 174-179, Dhaka, Bangladesh, 2009.

[16] V. Puente, C. Izu, J.A. Gregorio, R. Beivide, J. M. Prellezo and F. Vallejo, Improving parallel system performance by changing the arrangement of the network links, *Proceedings of the 14th International Conference on Supercomputing, 2000*, pp. 44 - 53, Santa Fe, New Mexico, USA, 2000.

[17] J.M. Kumar and L.M. Patnaik, "Extended Hypercube: A Hierarchical Interconnection Network of Hypercube," IEEE Trans. on Parallel and Distributed Systems, vol. 3, no. 1, pp. 45-57, 1992.

A CSMA Based Intra Cluster Communication Technique for Saving Cluster Head Energy

Amit Karmaker
M.Sc (Engg.) in Information & Communication Technology
Bangladesh University of Engineering & Technology
Dhaka, Bangladesh
amitkarmaker06@gmail.com

Md. Mahedee Hasan
M.Sc (Engg.) in Information & Communication Technology
Bangladesh University of Engineering & Technology
Dhaka, Bangladesh
mahedee.hasan@gmail.com

***Abstract*— Wireless sensor network is one of the most promising technology in wireless network. To design a sensor network, improving the lifetime of sensor node is critical issue. For these reason researchers in these field pay great attention to Medium Access Control (MAC) protocol. As most of the nodes are placed in remote and hazardous area, so when the power is down it is difficult to recharge them or it is not possible to replace them. For these reason the main concern of researchers is how to utilize the medium in a power effective manner. With advance of these various MAC protocols are introduced and lots of recent works are going on. LEACH protocols are one of these protocols which introduce clustering technique. In this paper, we propose a method for intra cluster communication based on CSMA and an extended cluster head selection algorithm. Also figure out mathematical comparison between LEACH protocol and our proposed method. The protocol selection depends on the application requirements and hardware characteristics. LEACH is suitable for low traffic and small area. Here we also consider low traffic.**

Keywords—Wireless Sensor Network; MAC protocol; Enhancing Lifetime.

I. INTRODUCTION

Wireless Sensor Network is the most emerging technology in the last decade during its robust application like environment monitoring, robotic exploration, military application and many others. With this advent of the rapid application, some critical situations are come into front. Power consumption is one of the most critical situations. Because, when the power of nodes are going down it is too much impractical to replace the node or power up of these node. Researchers in this area should keep in mind the energy dissipation factors.

Wei Ye et al.[2], llkerDemircol et al. [7] Identified some major sources of energy waste. Collision is first one. When a transmitted packet is corrupted it has to be discarded, and the follow-on retransmissions increase energy consumption. Collision increases latency as well. The second source is overhearing, meaning that a node picks up packets that are destined to other nodes. The third source is control packet overhead. Sending and receiving control packets consumes energy too, and less useful data packets can be transmitted. The last major source of inefficiency is idle listening, i.e., listening to receive possible traffic that is not sent. This is especially true in many sensor network applications. If nothing is sensed, nodes are in idle mode for most of the time. However, in many MAC protocols such as IEEE 802.11[1] or CDMA nodes must listen to the channel to receive possible traffic. Many measurements have shown that idle listening consumes 50–100% of the energy required for receiving. Different types of protocols are used to avoid that energy wastage. Most of them are basically application specific. LEACH protocol is one of the impressing protocols which use clustering method for saving energy. But it increase delay. In this paper we propose an intra-cluster communication method and average energy based cluster head selection technique for achieve better performance in context of both less energy consumption and less delay.

The remaining part of this paper is organized as follows: related work in Section II. In section III, problem statements are discussed. Section IV describes proposed methodology, section V describe the experimental result. Finally, section VI concludes the paper.

II. RELATED WORK

We said earlier that the main concern of most MAC protocol in the sensor network field is to reduce the energy consumed due to idle listening, overhearing, collision etc. To eliminate these drawbacks many protocols are drawn. Some protocols pay attention to reduce idle listening, where some other's attention to minimize latency. Some of them are contention based, some others are schedule based. IEEE 802.15.4 protocol combines both contention based and schedule based elements to achieve significant commercial impact [13]. S-MAC and T-MAC [3] use duty cycling method to reduce idle listening. S-MAC use fixed duty cycling, however it gives poor result for varying traffic load. T-MAC use dynamic duty cycling method over S-MAC to adapt it in various traffic load situation. But, it bottleneck by early sleeping problem. Both these protocol used Network Allocation Vector (NAV) to escape overhearing but it put forward idle active situation. Synchronization of duty cycle [11] also introduced to maximize sensor network performance. PAMAS [6] initiate a busy tone technique where other neighboring nodes need not to wait a NAV amount of time. When it hears busy tone, it permits to go to sleep mode. DW-MAC [10] is a low-overhead scheduling algorithm that allows nodes to wake up on demand during the Sleep period. AS-MAC [8] works over DW-MAC, it use Adaptive Scheduling

(AS) period into the operational cycle within which the nodes' active duration is made adaptive to variable traffic load. TRAMA [12] is a traffic adaptive protocol design whose main concentration is to reducing latency by using random access and schedule access technique. For doing these Venkatesh Rajendranet.et al. use three other techniques, neighbor Protocol (NP), Schedule Exchange Protocol (SEP) & Adaptive Election Algorithm (AEA). To perform all three techniques attractive CPU requirement must fulfill and it utilizes much memory compare to others.

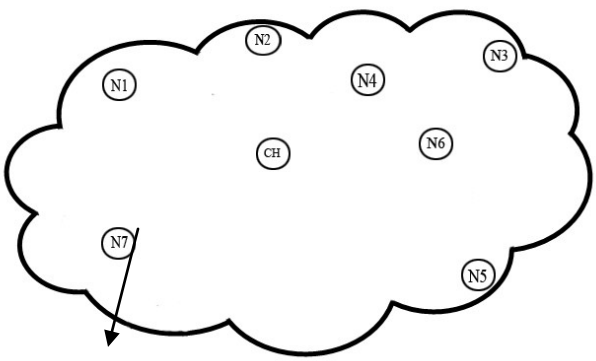
Later LEACH [9], an application specific protocol architecture developed. Which offer a cluster based self-adaptive and randomized technique to save power. In LEACH all nodes forming cluster and every cluster must have a cluster head node. These is responsible for collecting data from all non-cluster head node under its cluster using TDMA and then aggregate the data and send it to the base station (BS) using CDMA. Therefore, cluster head is more energy intensive than other node. So, if a node continuously remains as a cluster head node it decays its whole energy and after some round it falls. The LEACH incorporates randomized rotation of a high energy cluster head position such that it rotates among thesensors in order to avoid draining the battery of any one sensor in the network. In this procedure the power load is evenly distributed among the nodes.

III. Problem Statement

As we discuss in section I, we work over LEACH protocol and find out some problems.

A. Idle Listening by Cluster Head

We already discuss in the previous section that, LEACH is such kind of MAC which use clustering formula and intra-cluster communication between Cluster Head node and non-Cluster Head node is based on TDMA. The cluster head announce the time schedule for each non-cluster head node under its cluster. The cluster head remains awake during the whole time slot. In best case situation, it is definitely best approach. But in worst case situation where the traffic is lower enough that time it consume more energy.



Node 7 sense the medium and find something occur

Fig. 1. Single Cluster

B. High Latency

We identify another problem which can bottleneck the performance of LEACH protocol. Consider a situation where nodes are isolated from one another at a good distance

In figure (1), when node 7 senses the medium it finds something occur and it wants to transmit data to the cluster head. Whereas, cluster head previously announced the schedule of non-cluster head node. On this schedule node 7 gets its slot at very late period say after 30ms [figure (2)]. Previous 30 second of this round, Cluster Head remain idle because other than node 7, any node have not data to send. Thus it increases latency.

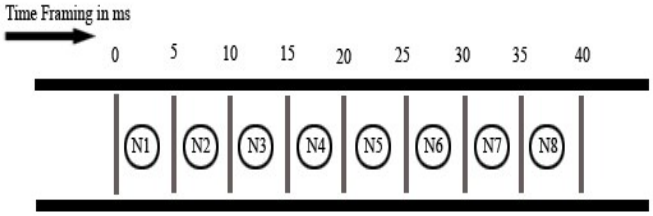


Fig. 2. Time Framing in TDMA

IV. Design of Proposed Model

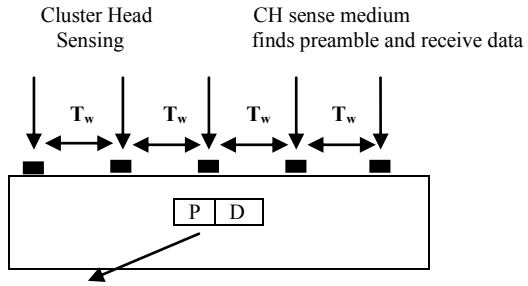
Our proposed method arrange into two parts. A CSMA based intra-cluster communication system followed by Enhancement of cluster head selection algorithm.

A. CSMA Based Intra-Cluster CommunicatioTechnique

Typically LEACH use TDMA for intra-cluster communication technique for transmitting data from non-cluster head node to cluster node. In section III, we discuss some problems by using TDMA. Now, here we initiate a CSMA based technique use preamble sampling. WiseMAC [5] and B-MAC [4] uses preamble sampling technique. WiseMAC, used dynamic preamble technique for downlink communication whereas B-MAC used fixed length preamble by using Clear channel Assessment (CCA). Like both of these MAC we sample the medium at fixed time length Tw. After Tw time the cluster head node awake and seek channel for data. If cluster-head found nothing it go to sleep mode. It senses the channel during its whole round. Now, it's a common question that what should be the size of preamble?

$$\text{Size of preamble} > \text{Two} * \text{wakeup Time} + \text{Sleep Time} \quad (1)$$

When two or more nodes want to transmit at the same time it first sense the channel in figure (3), if it finds the channel is occupied by other node it takes a random amount of time and then again sense. These procedure repeats until found the channel idle. Thus this technique is well suited for low traffic and where these nodes are placed for long term monitoring e.g. forest fire.



Nodes send data with preamble, if fails wait random wait of time

Fig. 3. Model of Our Proposed Method

B. *Enhancement of Existing Claster Head Selection Algorithm*

Assume that existing cluster head selection algorithm primarily select 10% cluster which we called Primary Selected Cluster Head (PSCH). Among them our proposed algorithm reduces 50% cluster head on the basis of average power. Since longevity is one of the main concern of LEACH protocol. So, we tried to keep equal power among all nodes. Enhancement of cluster head selection algorithm is given below.

Algorithm:

Step 1: Set a counter "CHCounter" for every senor, which determine the number of times it elected as cluster head.

Step 2: Calculate "avg_energy" (average energy).

Calculate avg_energy of the sensor node among the Primary Selected Cluster Head using the following formula

$$\text{avg_energy} = \sum_{i=0}^{n} \frac{E_i}{n} \qquad (2)$$

Here, n = number of Primary Selected Cluster Head (PSCH), Ei = Energy of each PSCH.

Step 3: Repeat step 4 and 5 until it reduces 50% of Primary selected cluster Head (PSCH).

Step 4: Pick a PSCH (Pick one after another. First, choose PSCH which have CHCounter = 0)

Step 5: Choose cluster head

If CHCounter = 0 and energy of the picked PSCH is greater than or equal of the avg_energy; select it as CH (Finally selected Cluster Head) and go to step 3

Otherwise if power of the picked PSCH is greater or equal of the avg_energy; select it as CH (Finally selected Cluster Head) and go to step 3

Step 6: Exit

Flowchart of above algorithm

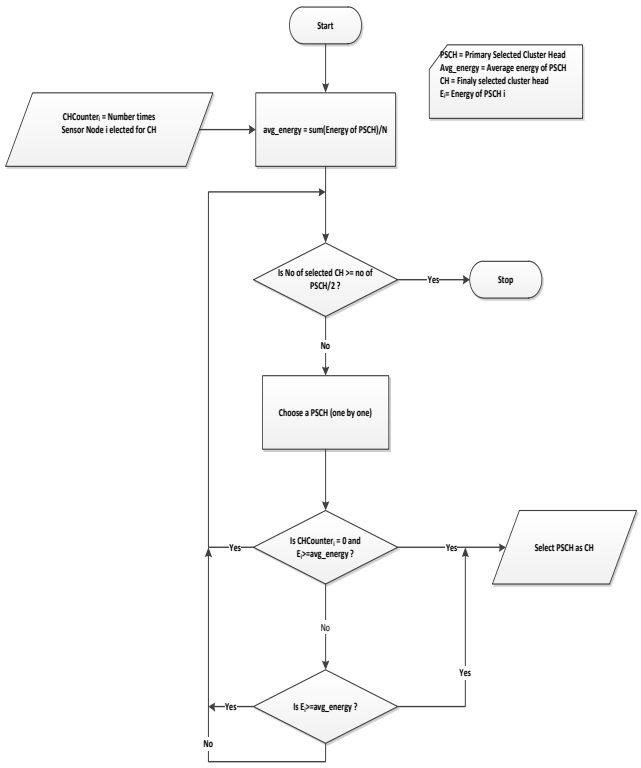


Fig. 4. Flow chart (Enhancement of exiting cluster head selection algorithm)

V. MATHEMATICAL ANALYSIS

In this section we show some mathematical calculation and simulation. Here we show the comparative analysis how cluster head consume energy of typical method and our proposed method. Also show the delay analysis between those two. We use MATLAB for our simulation.

In this analysis we only show the energy consumption by cluster head node and delay of intra cluster node. So, total area covered by cluster head node and how far the base station locates from the CH doesn't take into account. Table-I, shows the parameter which we take into account for our simulation.

TABLE I. SIMULATION PARAMETER

Parameter	Value
Node's Energy	500-600 Jules
Number of Nodes	50
Round Time	50 Seconds

Now, at first see the mathematical calculation. Let, taking 50 sensor nodes which mention in table I.

In traditional case, total wake up time of CH = Total Round time.

In Proposed case, total wake up time of CH = $\sum_{i=0}^{n}(Preamble\ Time + Data\ Processing\ Time)$

In Normal case its consumed energy = 50 Jules [Taking into consideration, per node consumed 1 Joule per round]

In Proposed case, if 25 nodes out of 50 nodes wants to transmit data to the CH,

Then, total consumed energy = 25 Jules [per node consumed 1 Joule per round] + 6.25 Jules [preamble consumed energy=.25 Jules]

= 31.25 Jules.

Here it saves 18.78 Jules in this current round. It saves much energy in low traffic.

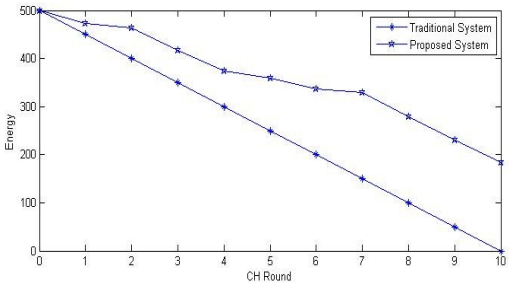


Fig. 5. CH Round Vs Energy

Figure-6 shows the discharging non cluster head node. Here, X-axis shows CH round and Y-axis shows Energy level. Figure shows that, normal case discharge its energy after 10 rounds [Only transmission energy taking into account, other factors are negligible, we avoid it. That's why curve look like linear]
Whereas, in our proposed system node contain more than 180 Jules after ten rounds.

On the other hand, Figure- 6 shows the comparative analysis of transmission delay of intra cluster nodes. X-axis contains the number of intra cluster node and Y-axis contains the Latency.

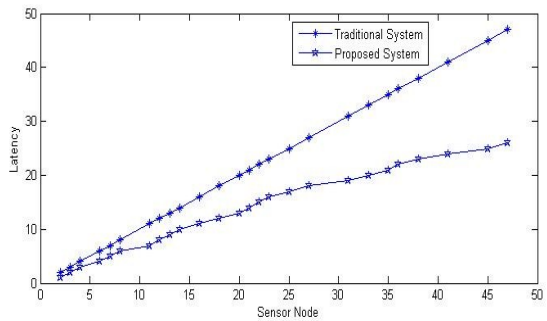


Fig. 6. Sensor Node Vs Latency

Figure-7 shows cluster head selection comparison between traditional model and our proposed model. It is shown that our proposed technique selects cluster heads which have energy above or equal to average energy. In traditional system, cluster head can be selected with below of average energy level. If it selects cluster head with below of energy level, its life time will be ended quickly comparatively others.

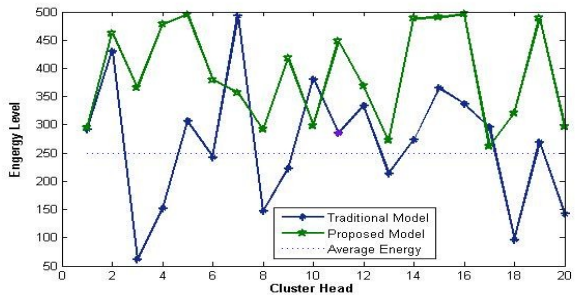


Fig. 7. Cluster head selection comparision

VI. Conclusion and Future Work

This paper proposed the enhancement of existing LEACH protocol. According to our simulation result, our proposed method gives better result in low traffic. It will decrease latency of sensor node to send data to the cluster head. Node need not wait for its time period to send data to cluster head.

Future work includes system scaling studies and parameter analysis. More tests will be done on larger test beds with different number of nodes.

References

[1] LANMAN Standards Committee of the IEEE Computer Society, Wireless LAN medium access control (MAC) and physical layer (PHY) specification, IEEE, New York, NY, USA, IEEE Std 802.11-1997 edition, 1997.

[2] Wei Ye, John Heidemann, Deborah Estrin, "An Energy-Efficient MAC Protocol for Wireless Sensor Networks" INFOCOM, volume 3, pages 1567-1576, June 2002.

[3] TijsvanDam, KoenLangendoen, "An Adaptive Energy-Efficient MAC Protocol for Wireless Sensor Networks", SenSys '03, pp. 171-180.

[4] JosephPolastre, JasonHill, DavidCuller, "Versatile Low Powe Media Access for Wireless Sensor Networks", SenSys '04 pp. 95-107

[5] A.El-Hoiydi, J,D.Decotignie, "WiseMAC: An Ultra Low Power MAC Protocol for the Downlink of Infrastructure Wireless Sensor Networks" Computers and Communications, 2004 ,ISCC,vol-1, pp. 244 – 251

[6] Suresh Singh, C.S. Raghavendra, "PAMAS - Power Aware Multi-Access protocol with Signalling for Ad Hoc Networks", Pages 5 -26

[7] Ilker Demirkol, Cem Ersoy, and Fatih Alagöz, Bogazici University, "MAC protocols for wireless sensor networks: a survey" IEEE Communications Magazine, vol. 44, no. 4, pp. 115-121, Apr. 2006.

[8] Y.Z.Zhao a, M.Ma b, C.Y.Miao a, T.N.Nguyen, "Anenergy-efficient and low-latency MAC protocol with Adaptive Scheduling for multi-hop wireless sensor networks", Pages 1452 – 1461

[9] M. J. Handy, M. Haase, D. Timmermann, "Low Energy Adaptive Clustering Hierarchy with Deterministic Cluster-Head Selection" Fourth IEEE Conference on Mobile and Wireless Communications Networks, Stockholm, erschienen in Proceedings, S. 368-372, ISBN 0-7803-7606-4World Scientific Publishing Co. Pte. Ltd., September 2002

[10] Yanjun Sun, Shu Du, Omer Gurewitz, David B. Johnson, "DW-MAC: A Low Latency, Energy Efficient Demand-Wakeup MAC Protocol for Wireless Sensor Networks" pp. 53-62 MobiHoc 2008.

[11] Y.Z.Zhao, C.Y.Miao, M.Ma, "An Energy-Efficient Self-Adaptive Duty Cycle MAC Protocol for Traffic-Dynamic Wireless Sensor Networks" February 2013, Volume 68, Issue 4, pp 1287-1315

[12] Venkatesh Rajendran, Katia Obraczka, J.J. Garcia-Luna-Aceves, "Energy-Efficient, Collision-Free Medium Access Controlfor Wireless Sensor Networks" Proc. ACM SenSys 03, Pages:181-192, Los Angeles, California, 5-7 November 2003

[13] Holger Karl, Andreas Willig, "Protocol and Architectures for Wireless Sensor Networks"John Wiley & Sons, 2005, pp.110-148.

Color Feature Based Video Content Extraction and Its Application for Poster Generation with Relevance Feedback

S.M. Mohidul Islam[1], Taznin Jahan[2], Banalata Das[3]
Computer Science and Engineering Discipline
Khulna University
Khulna, Bangladesh
[1]mohid_ku@yahoo.com, [2]jahantaznin@yahoo.com, [3]banalata.cse@gmail.com

Abstract— **Poster that represents a video has many applications in internet and media world. In this paper, a method is proposed to generate poster based on color feature and relevance feedback. Color feature is used because it is more closely to human perception and it is easy to extract, match and also effective for indexing. In our method as color feature of a frame in a video, HSV color space and Color Moment are used. To calculate similar frames i.e. to detect shot, Euclidean distance between the features of frames are compared with a predefined threshold value. One frame is selected from each shot and stored with Temporally Occurrence Value (*TOV*) of the Frame. As a video can have similar shots in it but in different times, so the similar frames from the similar shots of different times are detected and one frame is selected from them. Then this selected frame is stored with Permanent Occurrence Value (*POV*) of the Frame by replacing its *TOV* and discarding others of them from storage. Finally, five frames with maximum *POV* are selected to generate poster layout and ranked the posters by their relevance feedback. The experiments show the robust result than existing methods.***

Index Terms— **HSV; Color Space; Color Moment, Relevance Feedback; TOV; POV.**

I. Introduction

With the fast development of multimedia technology, a mass of digital videos have arisen. The content of video media is very rich, thus this increasing number of digital video is a great challenge to the management and maintenance of the digital video database. Browsing this vast volume of video data and expressing its contents has become much important. A video is a collection of video frames where each frame is a still image. When a video is being played, each frame is being displayed sequentially with a certain frame rate. With granularity from small to large, the segmentation results can be frame shot, scene and video. Shot is a sequence of frames recorded in a single-camera operation. Scene is a collection of consecutive shots that have semantic similarity in object, person, space and time. The typical frame rates are 30 and 25 frames/second as seen in the various video formats. An hour of video has 90,000-108,000 frames if it has a 30 or 25 frames/second rate, respectively [1]. No matter what kind of video format is used, this is a huge amount of data and it is inefficient to handle a video by using all the frames it has. A poster contains the most interesting and attractive parts of a video, while a summary sequence renders the impression of the content of an entire video. Among all types of video abstractions, poster represents the summary of a video that carries the highest semantic meaning of the content [1]. From a poster of a video, one can guess the overall video content. To generate poster layout of a video and/or to select a frame that represent a video, existing methods shows narrow results. Neither one is appropriate in terms of an expressive, effective and attractive video abstraction for the video content. Most video abstraction methods [3, 5] do not consider the structure information, such as the inter-shot relationship. In addition, there is a wide gap between the low level feature and video semantics for understanding the video content. The low level feature based approaches are inconsistent with human perception, and the semantics oriented methods are far from human expectation. C N Su, et al. [7] use Histogram in RGB color space to calculate the frame difference which cannot process appropriate result as color histogram based process has limited discriminating power [4]. Shot boundary is detected with a fixed length sliding window of current frames in most technique and these frames in each shot are clustered by the nearest neighboring rule in some process. Some

techniques [5] have been developed to detect a shot boundary automatically. Some technique use face detection scheme [3] and this process discover roles and their relationship is critical in movie content analysis. In this paper, color content and relevance feedback based video summarization technique i.e. poster generation technique is proposed. The experimental result shows robustness of the proposed technique in terms of expressiveness, effectiveness, attractiveness and time to summarize the video. The remainder of this paper is organized as follows. In Section II, the proposed method is described in details. In section III, we present some experimental results and conclusion is drawn in section IV.

II.PROPOSED METHOD

The proposed method is based on the color feature of video frames, as they are more closely to human perception and also as they are easy to extract, match and effective for indexing. For feature extraction we work on the HSV color model and calculate Color Moment, where Color Moment is represented by Skewness, Standard Deviation and Variance of each frame. Three features from HSV color model and three features from Color Moment are extracted. That is, total six features are used to represent a frame. To calculate similarity of two frames, we have used Euclidean distance between the features of two frames. Threshold (T) is used to compare the change between frame pairs. T is calculated as the average of Euclidean distances of all consecutive frames in a video. If the Euclidean distance of two consecutive frames is less than this threshold value then two frames are considered similar, otherwise they are different. Continuous Similar frames represent a shot of a video. We select a frame from each shot which represents the shot and also count the number of frames in that shot to estimate how long the shot is. We store the frame number of a shot with the selected frame. By this we have a storage containing a single frame from each shot and the number of total frames the shot consists of. As a video can have similar shots in it but in different time that means the similar shots are not consecutive, so in next step we select the similar shot of a video that are appeared in different time. If two shots are similar but are appeared in different times in video, we select one frame from the shots and add the number of frames of two shots together, i.e. one selected frame now represent many shots of a video those have similar features. To do this, we select a frame from the storage and compare it with all other frames in the storage. If two frames are similar in feature, then the value that represent total number of frames in shots are added and stored this updated value with selected frame and discarded other similar frames from storage. So now in storage we have only one frame from all similar shots of a video with the numbers of occurrence of frames in all similar shots of the video. Then we have selected first 5(five) frames from the storage with maximum value of occurrence to generate poster. Relevance feedback is considered for finalizing the most appropriate poster layout and will rank the poster in the storage according to the feedback. The system architecture of the proposed method can be figured as-

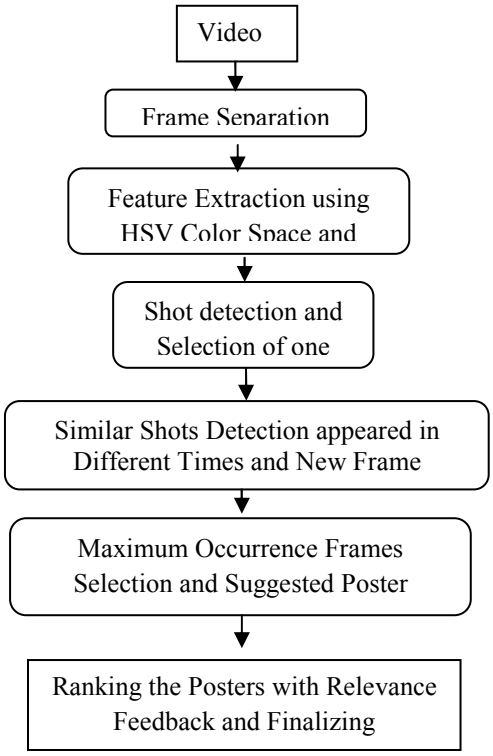


Fig 1: Proposed System Architecture

The System architecture is described in details in the following:

A. Frame Separation

We use Matlab toolkit to extract frames from video. It calculates total number of frames of a video and separates the frames in order. The process may be summarized as figure 2.

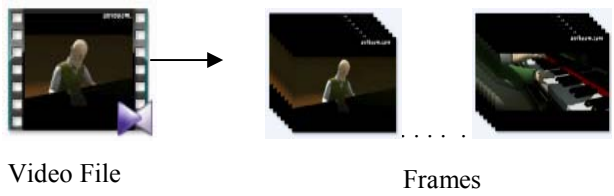


Fig 2: Video frame separation

B. Feature Extraction using HSV Color Space and Color Moment

For calculating feature vectors of a frame we have chosen color feature. Because color features are the most widely used visual features in image retrieval and they are easier to extract compared with other low level features [5]. To extract color feature, HSV (Hue, Saturation, Value) color space is used as they are closely with the categories of human color perception and approximately perceptually uniform and the RGB color space is not perceptually uniform. A color space is a multidimensional space in which the different dimensions represent the different components of color [2]. RGB color space of the frames is transformed into HSV color spaces using equation (1).

$$H = cos^{-1} \frac{1/2[(R-G)+(R-B)]}{\sqrt{(R-G)^2+(R-B)(G-B)}} \quad (1)$$

$$S = 1 - \frac{3[min(R,G,B)]}{R+G+B} \quad (2)$$

$$V = \frac{R+G+B}{3} \quad (3)$$

In equation (1) R, G, and B represent the red, green, and blue values of a frame respectively. Color Moment is derived from HSV color space. To represent Color Moment Variance, Standard Deviation and Skewness are used. That is, in proposed method six features (3 from HSV and 3 from Color Moment) are used to represent a frame. Variance is the average value of the quantity and can be defined by square of distance from mean, where mean denotes the average color value in the image. The variance (V_i) is represented as equation (1),(2),(3)

.

$$V_i = \sum_{N}^{j=1} \frac{1}{N} (P_{ij} - E_i)^2 \quad (4)$$

The Standard Deviation (σ_i) is the square root of the variance of the distribution, i.e.

$$\sigma_i = \sqrt{V_i} \quad (5)$$

Skewness (S_i) is a measure of the degree of asymmetry in the distribution and is

$$S_i = V_i \quad (6)$$

In above equations, E_i is mean and p_{ij} is i^{th} color channel at the jth pixel of a frame[2]. Figure (3) represents this step

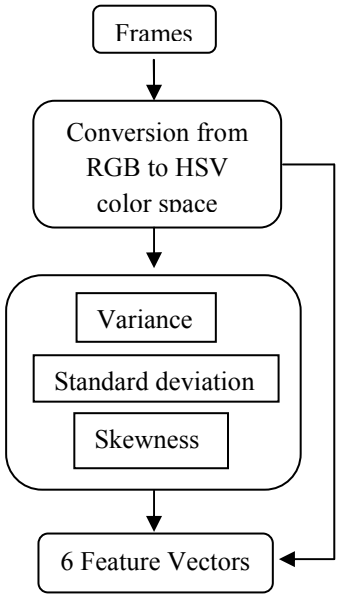
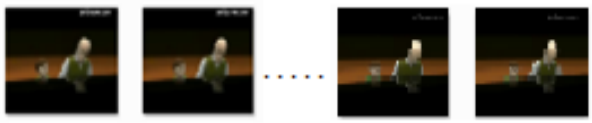


Fig 3: Features extraction process

C. Shot Detection and Selection of one Frame from each Shot

Shot is a sequence of frames with similar features. To detect similar frames, we used a threshold value (T) which is calculated as the average value of Euclidean distances of any feature of all consecutive frames in the video. That is, different threshold values are used for different features. The Euclidean distance is measured as Di= $\sqrt{\sum(xi - yi)^2}$ where x_i and y_i represent the feature of two consecutive frames. To detect the similar frames, the Euclidean distance of two consecutive frames is calculated and compared this value with threshold T. If the value of distance for two frames is less than the threshold value then the consecutive frames are considered as similar frames i.e. they have similar semantics otherwise they are different. To detect shot, this distance rule of consecutive frames is applied. All the frames are traced until a different frame is found and this collection of frames represents a shot. Then one frame is selected from the shot and stores it in storage with the total number of frames that shot consist of. This value of total number of frames is called Temporally Occurrence Value (*TOV*) of a Frame. The process is repeated to detect all shots and one frame is selected from each shot and stores it in the storage with their corresponding *TOV*. For example, Fig 4(a) represents the frames of a shot and fig 4(b) represents the selected frame from the shot



(a)

(b)

Figure 4: (a) Frames of a Shot and (b) selected frame from the shot

The total number of frames in this shot = 142. So for the representative frame of this shot, *TOV* value = 142 and the *TOV* value is stored in the storage with the representative frame.

D. Similar Shots Detection appeared in Different Times and New Frame Selection with Occurrence

After getting a TOV from each shot we have to calculate Permanent Occurrence Value (*POV*) of a frame by comparing each frame with others in the storage. In this case, the Euclidean distance of each frame from all other frames is calculated and each case the distance is compared with the previous threshold value *T* (different *T* for different features). If the distance of two frames is less than T, then these frames are similar and considered they represent different time shots of video with similar features. So from these similar frames of shots with similar feature, only one frame is selected and other is deleted from the storage with their *TOV* and the *TOV* values of all the frames are added together to construct *POV*. This *POV* value with the selected frame is stored in the storage. This removes the repetition of representative frame of similar shots in the video.

For example, consider the two representative frames of a video that are similar in figure 5.

(a) (b)

Figure 5: Two frames with similar features

TOV value of Frame of figure 5(a) =142 and *TOV* value of Frame of figure 5(b) =114. These two representative frames are found similar, so one frame is selected with *POV* value = 142 + 114 = 256 and is replaced its *TOV* by its *POV* value in the storage and other frame with its *TOV* value is discarded from the storage.

E. Maximum Occurrence Frames Selection and Suggested Poster Layout Generation

The representative frames of the storage are ordered by their *POV* values and selected five maximum occurrence frames to generate suggested poster layout. Five Posters are generated. One poster is generated using only one frame that has largest *POV* value, another poster is generated with two frames of consecutive largest *POV* values, another with frames with three largest *POV* values and so on.

F. Ranking the Posters with Relevance Feedback and Finalizing

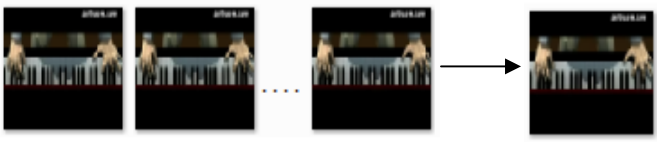
The poster is finalized from relevance feedback. The posters are ranked with relevance feedback value and is selected the final poster from their rank. The poster with highest rank will be select as final poster.

III. EXPERIMENTAL RESULTS

We used several videos for our experiment. In the following the result for video 'Piano.avi' is shown. After separating the frames from 'Piano.avi' video, the shots and the frames are selected as follows-

Shots Selected frame

Total number of frames in this shot, i.e. *TOV* = 130

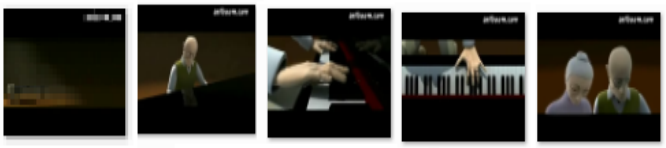


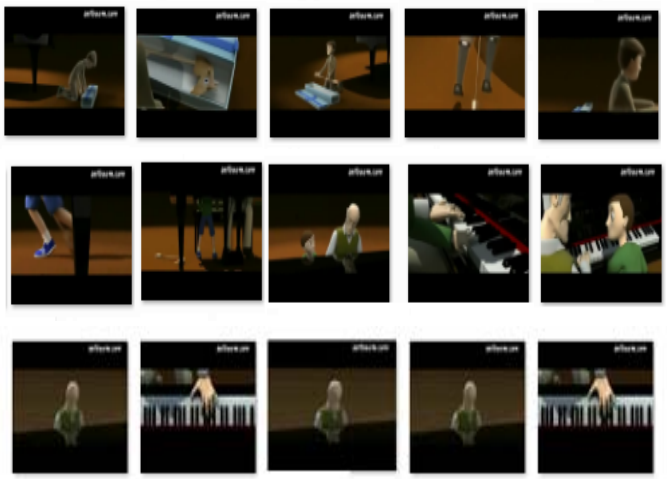
Total number of frames in this shot, i.e. *TOV*= 301

⋮

Figure 6: Shots and Selected frame from each shot with *TOV* value

A part of several selected frames from the shots of the video are shown in figure 7.





. . . . . .

Figure 7: Several representative frames of video 'Piano.avi'

Figure 8 shows the ordering of representative frames by value of *POV*. Mention that, all selected frames are not shown in Figure 6, 7 and 8.

Figure 8: Ordered Representataive frames of video 'Piano.avi'

Where *POV* for first frame = 304, for second frame = 278, for third frame = 204, for fourth frame = 173, for fifth frame = 181, and so on. The poster layout is generated from frames with maximum *POV* as figure 9.

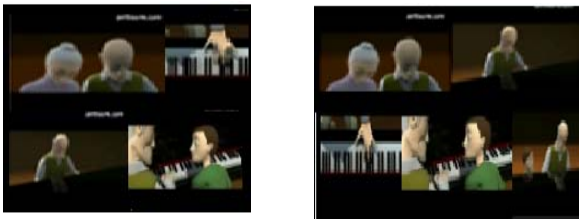


Figure 9: Five Suggested Poster layouts

From relevance feedback poster layouts can be updated. As example if any user wants to generate a poster using only one frame that is the second frame in figure 8. Then from users relevance feedback poster layouts are ranked and finalized. In figure 10, the poster with their ranks and relevance feedback is shown. That mean the poster 3 and 5 are more acceptable to user.

Rank 1: Poster 4 (90%) | Rank 2: Poster 3 (80%) | Rank 3: Poster 5 (75%)

Rank 4: Poster 2 (25%) | Rank 5: Poster 1 (15%)

Figure 10: Ranking the poster according to Relevance Feedback

IV. CONCLUSION

In this paper, we have introduced a robust single frame selection process from similar shots of a video with color feature extraction techniques and six different features are used for selecting the shot and frames from the video to generate poster. The frequency of appearance of a shot or representative frames for that shot is considered also to suggest poster layout. Relevance Feedback is used to rank and finalize the most appropriate poster. Result shows the real impact of video poster effectively. This line of work has the

potential to greatly enhance video labeling tasks which are increasingly of interest for activity recognition, poster generation and other applications of video summarization. The future work is doing the feature extraction of the image with more robust technique and with different distance measure for better result.

REFERENCES

[1] Jung Hwan Oh, Quan Wen, Sae Hwang, Jeongkyu Lee, "Video Abstraction", *The University of Texas at Arlington,* USA.

[2] S. Mangijao Singh, K. Hemachandran, "Content-Based Image Retrieval using Color Moment and Gabor Texture Feature", *IJCSI International Journal of Computer Science Issues*, Vol. 9, Issue 5, No 1, September 2012 , ISSN (Online): 1694-0814.

[3] Yan Wang, Tao Mei1, Xian-Sheng Hua, "Community Discovery from Movie and Its Application to Poster Generation", *17th International Multimedia Modeling Conference,* MMM 2011 Taipei, Taiwan, January 5-7, 2011 Proceedings, Part I.

[4] Y Wang, "Research on content-based news video abstraction technology "Subject: Signal and Information Processing", *GTID:2198330332478667,* 2011-06-10.

[5]M. Omidyeganeh, S. Ghaemmaghami, S. Shirmohammadi ,"An Event Based Approach to Video Analysis and Keyframe Selection", *Electrical Engineering Department, Sharif University of Technology , Distributed and Collaborative Virtual Environment Research Lab.,* University of Ottawa.

[6] Annesha Malakar, Joydeep Mukherjee, "Image Clustering using Color Moments, Histogram", *Edge and K-means Clustering, International Journal of Science and Research (IJSR*), India Online ISSN: 2319-7064.

[7] C N Su, "Video summary of research and systematic implementation ", *GTID2208360278970231,* 2010-09-17.

Cloud-Niagara: A High Availability and Low Overhead Fault Tolerance Middleware for the Cloud

Asif Imran*, Alim Ul Gias, Rayhanur Rahman, Amit Seal, Tajkia Rahman, Farhan Ishraque, and Kazi Sakib
Email: asif.imran.anik@gmail.com, alimulgias@gmail.com,
bit0101, bit0122, bit0107, bit0201@iit.du.ac.bd, sakib@univdhaka.edu
Institute of Information Technology, University of Dhaka

***Abstract*—Fault tolerance is the ability to a system to continue its functionality despite the presence of faults in the architecture. For a dynamic system such as the cloud, fault tolerance is required to ensure business continuity. This paper proposes a high availability middleware that ensures fault tolerance for cloud based applications. Effective Descriptive Set Theory is used to determine the model of fault detection for real life applications running on the open source cloud. A deterministic algorithm of the middleware is provided that achieves automatic allocation of backup nodes to the system based on the faults. After detection of faults, the middleware directs the system to add new nodes as replicas of the failed nodes, ensuring continuity of the cloud applications. Next, a case study including seven real life applications such as PostGreSQL Database, etc are described and fault tolerance is ensured through the proposed middleware. Empirical performance analysis of the algorithm is carried out and results are compared to traditional systems.Results show that in the presence of faults induced during experimentation, the middleware can be effectively used to introduce replica and ensure fault tolerance of bottleneck resources for executing 700 to 1000 processes per unit time.**

I. Introduction

Fault tolerance is the ability of a system to perform its function properly in case of faults and excessive system loads. Cloud computing is the dynamic provisioning of virtual machine instances from a shared resource pool of physical machines. It provides dynamically distributed entities for rapid scaling and dynamic provisioning to increase performance. For this reason, fault tolerance needs to be ensured for cloud services to prevent lower level errors from propagating into system failure. Ensuring fault tolerance for a largely dynamic system such as the cloud presents significant research challenges, since a large number of processes in different states are involved. Attributes of the algorithm such as watchdogging, checkpointing and logging critical processes have not been considered for the cloud. Modest studies have been conducted to ensure fault tolerance in traditional systems [1], [2]. More specifically, the research issues that need to be addressed are provided in the following.

1) Mathematically analyze causes of faults in the cloud. Using the analyses, propose an algorithm for a middleware to ensure high availability fault tolerance of the cloud platform through rapid checkpointing, watchdogging and logging.
2) Empirical investigation to analyse the performance of the proposed algorithm of the middleware in a real life case scenario.

A modest number of reports and papers have emanated from the study to date, including a general overview [1], papers concerned with protection of distributed systems from failures were highlighted in [2], [3]. Preliminary account on fault detection and recovery through replication is presented in [4]. However, providing an algorithm for a middleware to ensure high availability fault tolerance have been considered to a limited extent.

Based on the research issues stated above, this paper proposes an algorithm of a middleware that ensures fault tolerance of cloud. The algorithm of the middleware is deterministic and automatically checkpoints and logs system faults. Upon detection of faults, the algorithm directs the system to provision replicas to faulty machines. The proposed middleware is called Cloud-Niagara because it is capable of ensuring fault tolerance at excessive system loads and existing faults through incorporating multiple primary controllers and using dynamic group architecture. The benchmark for industry level Central Processing Unit (CPU) and Random Access Memory (RAM) is considered to be 70% [5]. Effective Descriptive Set Theory is used to model the occurrence of faults in the cloud [6].

In the case study, seven real life applications namely Online Survey application (OSa), Source Code Repository (Rep) applications, Open Source Project Manage-

978-1-4799-3497-3/13/$31.00 ©2013 IEEE

ment application called HTProjects (Proj), PostgreSQL Database (DB) Management Applications, File Transfer Protocol (FTP) services, Nova Dashboard (ND) for OpenStack Compute Cloud and Domain Name Services (DNS) were executed in the cloud. Next, Cloud-Niagara was applied in the distributed controllers and used to monitor the resource consumption of the applications. Additional instances were placed to act as replicas to achieve fault tolerance. The results obtained from the case study are used to show how the middleware performs in achieving fault tolerance in the cloud. The results reflect system load exceeding 70% when no replica or fault tolerance mechanism is in place for executing 700 to 1000 processes concurrently per unit time.

II. Related Work

Preventing failures in systems through detection of peer failures and obtaining lost information in large scale Distributed Hash Tables is proposed in [1]. Horizontal Active Replication scheme is presented that ensures viewing key processes to detect failures. The mechanism ensures an effective range of query processing, even when probability of failure is significantly high. However, load balancing and failure detection is carried out by a master load server that may lead to low availability.

Linear Finite State Machines (LFSMs) have been made fault tolerant in discrete time dynamic systems, such as simulations [2]. Redundant implementations are used to construct fault tolerant systems, that involve separate machines to preserve state, evolution and properties of target systems. The redundant machine protect the data of the target nodes through encoding. However, the communication between the remote machines that ensure redundancy result in high latency of the architecture.

Automatic addition of new fault handlers into an existing fault tolerance scheme is proposed in [3], [4]. The assumption is based on the problem of anticipating faults of all program classes at the time of system design. The methodology involved updating new faults at the system design level through incorporating of a random fault generating module. Significant contribution of effective resource provisioning have been made in [7]. However the performance benchmarks of resource utilization were not considered.

Fault tolerance is achievable means of active redundancy as proposed in [5]. Dependability analysis needs to be carried out to analyze the degree of redundancy. Self-checkers are installed that is used to differentiate the faults. The self-checker is required to be intrinsic and fail-stop activities of the component based distributed system needs to be placed. However, each message in the methodology is required to have multiple components executed physically which results in increased traffic.

Protocols like Isis, Amoeba and Totem deliver messages similarly in the total order [8]. Those protocols are widely used to maintain consistency in replication of fault tolerant systems [9]. However, delays are introduced when atomic multicast potocols are used. Delays in messaging protocols can be prevented if the primary is responsible for deciding the ordering of the operations. However, ensuring the high availability of the primary node through a middleware is a significant challenge. The high number of message passing in synchronous communication will be very costly for the consensus algorithm.

III. Fault Standard of Cloud-Niagara

As discussed in the previous section, ensuring high availability fault tolerance requires a middleware that can monitor resource usage and send notifications without delay. The proposed architecture of Cloud-Niagara satisfies the requirement discussed here. Cloud application failures arise from incorrect program designs or coding errors but, more often than not, they arise from transient and nondeterministic errors. A mathematical model needs to be provided to analyse the type of faults.

A. Effective Descriptive Set for Fault Identification

The system S_n consists a finite set of real life application A_i where $A_i = \left(a_1, \quad a_2, \quad \cdots \quad , a_u\right)$. Each application a_i consists of a finite set of processes P_j denoting set of all processes such that the condition $\forall a \in A : P_j = \left(p_1, \quad p_2, \quad \cdots \quad , p_v\right)$ is satisfied where u and v are positive integers. Each process p_c have states that are represented by a finite set $S_t = \left(s_1, \quad s_2, \quad \cdots \quad , s_n\right)$. The states are divided into two subsets representing the correct and failed sets respectively, where the value is obtained by $S_{correct} = \left(s_{l1}, \quad s_{l2}, \quad \cdots \quad , s_{ln}\right)$ and incorrect state as $S_{failed} = \left(s\prime_{l1}, \quad s\prime_{l2}, \quad \cdots \quad , s\prime_{ln}\right)$. Hence the above cases are true if and only if

$$(S_{correct} \bigcup S_{failed} \in S_t) \wedge (S_{correct} \bigcap S_{failed} \in \emptyset)$$

are true. The combined states of the various processes in the real life applications running on the cloud platform is satisfied by the following equation 1.

$$\begin{aligned} B_i = (\forall i \in 0 < i \leq n : b_i \subset S_{correct}) \wedge \\ (\forall j \in 0 < j \leq n : b_j \subset S_{failed}) \end{aligned} \tag{1}$$

Valid write counts is projected as $(s_l, s\prime_l) : S_t \in \delta_p \wedge s_l, s\prime_l \in S_t$. Here δ_p is the probability of a valid state. Write operations are valid if and only if a process p_j executes operation w_j on the file that is allowed access.

Hence, if $w_j\prime$ is a set of write operations that result in process failure, we get,

$$w_j\prime = (s_l, s_l\prime) : (\exists p_j : x \notin w[j] \wedge x = S_t) \quad (2)$$

Read operations can be carried out in multiple files. Hence the files that are in specific group denoted by $G(R_k)$ is described here.

$$\begin{aligned} G(R_k)(s_l, s_l\prime) &= ((s_l\prime) : \forall y : y \in r_y : v(s_l) \\ \neq v(s_l\prime) \vee v(s_l\prime) &= v(s_l) \vee \forall y : y \notin r_y : v(s_l) \\ &= v(s_l\prime) \wedge v(s_l\prime) \neq v(s_l)) \end{aligned} \quad (3)$$

An application a_i $(0 < i \leq u)$ consists of a process p_j $(0 < j \leq v)$ that carries out read and write operations on the data stored in the cloud. To ensure low latency fault detection by proposed architecture, faults must be detected with minimal delay in which those occur [3]. Read operations are represented by the set R_x and write operations are represented by the set W_y. After read or write operation by p_j, the state is saved in a state variable s_l and updated in S_t. If the state is faulty the $S_t : s_l = s\prime_l$ else $S_t : s_l$. Hence, the conditions need to satisfied if s_l is to be correct and not failed,

$$\begin{aligned} &\exists p \in P_j : G(s_l) \subseteq S_t \vee G(s_l) \subseteq S\prime_t) \\ \implies &G(s_l) \subseteq R_x \wedge G(s_l) \subseteq R\prime_x = \emptyset \vee \\ &G(s_l) \subseteq W_y \wedge G(s_l) \subseteq W\prime_y = \emptyset; \end{aligned} \quad (4)$$

IV. Proposed Algorithm for Cloud-Niagara

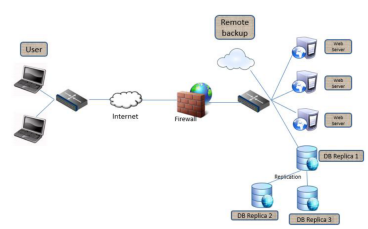
High availability fault tolerance is ensured by Cloud-Niagara through resource utilization monitoring and notifying the primary controller. The Algorithm 1 presents the algorithm of Cloud-Niagara and identifies the input and output of the distributed architecture. The algorithm notifies the system to add new nodes once the utilization of resources exceeds the predefined threshold. N_{node} is a queue that holds the node id of each node that is allocated to work with the cloud system. For each entry in N_{node}, the ratio of $R_{current}$ that is currently used resource of a node is calculated. The total resource available of that node is denoted by R_{actual} that is calculated simultaneously. Here M is the maximum number of resources available and $N_{temporary}$ are the additional resources added for a specific time period t.

The obtained value of R_{actual} is checked against upper threshold value T_{upper}. If the ratio is more than or equal to the threshold value, the node identity is added to another queue N_{exceed}. Else it is added to another queue N_{low}. Finally these queues are sorted in ascending order of node identities.The following section highlights a real life case study used for testing the proposed algorithm of Cloud-Niagara. Out of the total number of processes, let H be the number of processes that are faulty at time

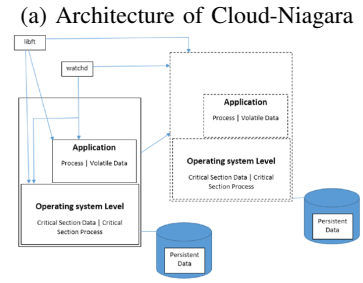
Algorithm 1: Algorithm of Cloud-Niagara

Input: Resource Usage data from the slave nodes $\langle a_1, a_2, \ldots, a_n \rangle$
Output: Graphical of output of resource usage information and warning about faulty nodes

```
location ← 0
i ← 2
while n ← N_nodes do
    if R_current/R_actual ≥ T_upper then N_exceed ← N_exceed + 1;
    else R_current/R_actual < T_lower N_low ← N_low + 1;
Record and sort all indexes J in ascending order of R_current/R_actual;
if N_exceed == N_node then
    Generate Warning for admin for adding more physical resource;
    N_node = N_node − 1;
    Add new node to Cloud with back up id I;
    J[i] = J[i] + 1;

if N_now ≥ 2 then
    Set m equal first index in J;

if M == 0 then
    N_node = N_node − N_m;
    Res[i] = Res[i] − Res[j]
    N_node = N_node − 1;
    N − low = N_low − 1 and J[i] = J[i-m];

if N_low ≥ 2 then
    n = J[1];
    N_node = N_node − N_temporary;
    N_node = N_node − 1;
    BackID ← 0

while x ← N_nodes do
    N_x = BackID;
return Graphical output of resource usage at different nodes
```



(a) Architecture of Cloud-Niagara



(b) Watchdog, Checkpoint and Log mechanisms of Cloud-Niagara

Fig. 1: Fault Tolerant Mechanism of the proposed solution

T that eventually leads to failure of the cloud service, hence $\lceil \frac{N}{2} \rceil \leq H \leq \lceil N \rceil$ is satisfied where N is the total number of nodes. Initially, no system has failed N nodes function correctly. Let a failed state satisfying the condition H is denoted by F.

$$\Delta_s = (N - q) * 0.30 \tag{5}$$

$$\Delta_s = (N - q) * 0.30 + \Delta_{primary} \tag{6}$$

The system can be in complete functional state when N nodes are active. Next, we have a condition where q nodes satisfy H. We see that (N-q) nodes are still active where $(0 \leq q \leq F_{max})$, where F_{max} are the maximum number of failed nodes. W define an acceptable state of machine as $(0 \leq q \leq 0.30N)$. Therefore, $F_{max} \leq 0.30N$ condition is required to be satisfied for the system to function despite system faults. The coverage factor c is defined as 0.30 since the benchmark considered for fault tolerance is 70%. The Δ_s is defined in equation 4 and Δ_s when main node fails is denoted by equation 5. The failure detection is captured by Cloud-Niagara using the mathematical model described here. The functionality of the model ensures high availability and low latency of the fault tolerance scheme. The time taken to recover from the failure is required to be compared to identify the performance of Cloud-Niagara compared to traditional mechanisms.

To compare performance of Cloud-Niagara with existing architectures, Mean Time To Recover (MTTR) from the failed state back to functionality is compared.

$$\begin{aligned} MTTR_q &= \frac{\sum_{q=1}^{n}(q)(dT)^2}{q} \\ &= \frac{\sum_{q=1}^{n}(q)(T_{recover} - T_{failure})^2}{q} \end{aligned} \tag{7}$$

The time for failure is the product of coverage factor c and number of failed nodes H. Hence

$$T = c \times H : \left(h_1, \quad h_2, \quad \cdots \quad , h_n\right) \leq 0.30N \tag{8}$$

Cloud-Niagara uses active replication technique to make the system fault tolerant. This technique gives all the replicas the same role unlike the primary-backup technique [10]. A client process pi invokes operation op_{arg} to one replicated server x. All the replicas of x will receive the invocation op_{arg}. Next, all of them will send the response to the p_i process. The process can function if replicas do not behave maliciously.

V. Case Study of Real Life Applications for Fault Tolerance

The performance of Cloud-Niagara is evaluated through a case study on real life Openstack cloud. The middle-ware is installed in the cloud controller.

```
Cloud-Niagara: info: [2874] dispatching hogs: 0 cpu, 0 io, 7 vm, 0 hdd
Cloud-Niagara: FAIL: [2880] (495) hogvm malloc failed: Cannot allocate memory
Cloud-Niagara: FAIL: [2874] (395) <-- worker 2880 returned error 1
Cloud-Niagara: WARN: [2874] (397) now reaping child worker processes
Cloud-Niagara: FAIL: [2874] (452) failed run completed in 1s
```

Fig. 2: Resource Overloading fault warnings obtained from Cloud-Niagara

A. Scenario

The cloud infrastructure has 10 physical nodes and a controller. The middleware monitors resource usage for all the physical nodes in the cloud. The processes include running seven services discussed previously that are commonly used in real life [11].

The experiments were carried out in the cloud infrastructure of a real life cloud service provider using OpenStack Essex Cloud on Ubuntu 12.04 Long Term Service (LTS) server. The experiments in the case study included running the applications in virtual machine instances of the cloud [12]. Each instance is allocated 2 GigaBytes of Random Access Memory (RAM) and 2 Central Processing Unit (CPU) cores. The RAM and CPU are increased or decreased based on the information obtained from the proposed algorithm.

B. Bottleneck Resources

Forking processes are executed on the virtual machine instances of the cloud to load the CPU and RAM. As part of the assumptions, we also consider that the cloud infrastructure can be scaled both vertically and horizontally. The Cloud-Niagara middleware operates on the cloud controller that acts as the primary.

Cloud-Niagara monitors the system utilization and provides the resource consumption in graphical format. Once the memory allocation or CPU consumption by each core exceeds 70%, failing and warning messages are exhibited by Cloud-Niagara as shown in the Figure 2. Bandwidth is not a bottleneck when it comes to the scenario of the case study [13].

C. Methodology to Add and Remove Fault Tolerant Resources

Once a backup is there in place it needs to be added into the system. Cloud-Niagara manages addition of new backups in the following format.

1) Introducing new backup resource using Cloud-Niagara: The backup machine sends a message using multicast protocol to the internally connected group. The contents of the message involves a Backup-ID. Next the controller that acts as the primary will make a priority list in a queue and send the list over the internal network.

The Backup-ID is used to notify the accepted backup machine that responds with an acknowledgement. Then the new backup is integrated that begins to update itself from the latest check-pointed state. The checkpoint and watchdog daemons play critical roles in this regard [14]. Once the new membership is acknowledged, all processes are sent a notification message about the new core or memory unit and mapped to a specific node.

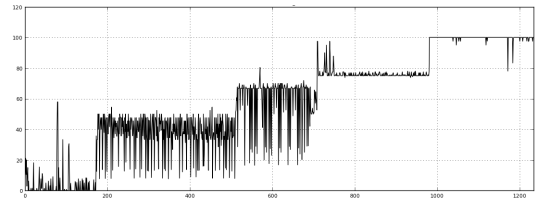
2) Removing existing backup using Cloud-Niagara: As a node becomes faulty it must be replaced. The backup removal procedure of Cloud-Niagara is replaced using broadcast messages. The primary node issues a message to remove a specific backup with a unique Backup-ID [15]. This message is broadcasted in the internal cloud network [16]. The nodes can be removed as a high priority machine requires more resources.

D. Analysis of Obtained Results

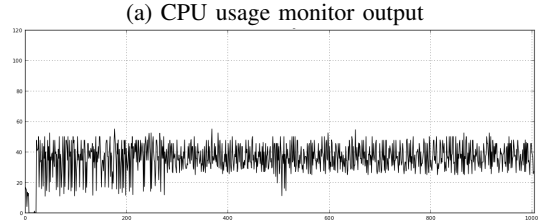
As replicas are activated, it is seen that the resources are capable of taking in increased CPU and Memory loads as depicted in Figure 3a and Figure 3b. The accumulated results are shown in Table 1 and graphs denote the number of processes triggered per unit time on the horizontal axis. The vertical axis denotes the percentage of resource utilization for the specified number of CPU cores and RAM. The figures show that the systems gradual memory and CPU consumption can be kept below the benchmark level by adding more replicas. The ring architecture of the fault tolerant nodes contributes to fault tolerance of the cloud service [11].

Initially the CPU load is under the danger level. Figure 3a shows the gradual growth of CPU utilization as the backup processors are added in accordance to the demand of the primary CPU. When no replica is used, an increased load on the CPU will result in the single Core being continuously used to full potential [17]. Increase of load at 100 units results in a sudden increase of resource utilization in Figure 3a. When the load per unit time is increased to 180 units, the CPU usage is normal. The situation remains the same till 520 processes executed per unit time. During this period, the CPU utilization of the instances remains at around the 50% scale because of provisioning by Cloud-Niagara. From 520 processes to 700 processes, the CPU load is seen to increase from 50% to 70% benchmark with a sudden increase to 80% for 580 processes per unit time.

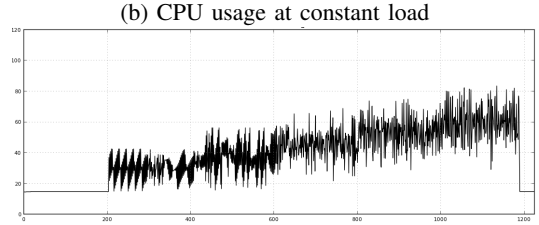
The graph in Figure 3b calculate average CPU utilization for FTP and DNS are shown in Figure 3b. For the two services, for 20 processes executed per unit time, the CPU utilization declined momentarily. This is caused by the sudden decrease in DNS and FTP processes as no user required to establish an FTP or execute DNS for that time period. From 20 processes per unit time onwards the resource usage increased to 50% and remained at the desired scale of 30% to 50% that is below the prespecified threat level of 70%. The cloud infrastructure sustains the load for upto 1000 processes executed per unit time.



(a) CPU usage monitor output



(b) CPU usage at constant load



(c) Graph of memory utilization monitored through Cloud-Niagara

Fig. 3: Results obtained for Cloud-Niagara

Constant increase in RAM usage is identified by Cloud-Niagara and represented graphically as shown in . The graph in Figure 3c highlights the memory utilization as increased load is applied. The system increases load significantly from 0 to 1200 processes per unit time. Initially, the utilization of RAM is negligible for the 200 processes. This is because the first 200 processes are CPU intensive and consumed memory of upto 18% only. At the load of 1000 forking processes, the system load reaches 80% to 90%, well above the threshold level of 70%. This shows that the current physical memory will remain in the danger level above 70% if the processes executed per unit time ranges from 900 to 1200. Such load will cause the system to become more prone to failure.

The important information derived from Cloud-Niagara can be used to decide when to add new resources to the existing cloud infrastructure. Hence the cloud architecture can be made fault tolerant using Cloud-

Service	Cloud VM	Replica Available	Evaluation Time	Throughput	Num.of Fault	Allocate Replica
OSa	1	1	500X9	50	17	No
Rep	2	1	1000X5	67	19	No
Proj	1	1	700X7	96	12	Yes
DB	4	2	2000X6	87	20	Yes
FTP	1	1	600X7	94	12	Yes
ND	3	2	600X4	48	12	No
DNS	2	1	400X8	76	27	No

TABLE I: Obtained results of fault detection by Cloud-Niagara

Niagara's algorithm.

VI. Conclusion and Future Work

This paper aims to provide a high availability and low latency fault tolerance middleware for the cloud. The paper presents a mathematical model based on Effective Descriptive Set Theory for fault detection. In addition, the replication standards and empirical models for performance analysis of Cloud-Niagara is provided. Next, a deterministic algorithm based on watchdogging, checkpointing and journaling has been discussed for the middleware. Dynamic group communication of the algorithm in terms of read and write operations ensure effective fault tolerance.

Real life case study is shown that reflects the performance of Cloud-Niagara middle-ware in replication of CPU and Memory units. Overall utilization of the cloud service is kept below the pre-specified benchmark and resources are made more fault tolerant through a component based architecture. The results reflect that resource utilization can be effectively monitored using Cloud-Niagara and replica can be added or removed according to the need of the primary. The empirical cost estimation of Cloud-Niagara in terms of scalability and performance is a subject of future research interest.

VII. Acknowledgment

This research has been supported by the University Grants Commission, Bangladesh under the Dhaka University Teachers Research Grant No-Regi/Admn-3/2012-2013/13190.

References

[1] R. Natella, D. Cotroneo, J. Duraes, and H. Madeira, "On fault representativeness of software fault injection," *Software Engineering, IEEE Transactions on*, vol. 24, no. 7, pp. 1313–1327, 2013.

[2] T. Pitoura, N. Ntarmos, and P. Triantafillou, "Saturn: range queries, load balancing and fault tolerance in dht data systems," *Knowledge and Data Engineering, IEEE Transactions on*, vol. 24, no. 7, pp. 1213–1229, 2012.

[3] R.-T. Wang, "A dependent model for fault tolerant software systems during debugging," *Reliability, IEEE Transactions on*, vol. 61, no. 2, pp. 504–515, 2012.

[4] W. Xu and T. Zhang, "A time-aware fault tolerance scheme to improve reliability of multilevel phase-change memory in the presence of significant resistance drift," *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on*, vol. 19, no. 8, pp. 1357–1367, 2011.

[5] B. Littlewood and J. Rushby, "Reasoning about the reliability of diverse two-channel systems in which one channel is," *Software Engineering, IEEE Transactions on*, vol. 24, no. 7, pp. 1313–1327, 2011.

[6] H. Becker and A. S. Kechris, *The descriptive set theory of Polish group actions*. Cambridge University Press, 1996, vol. 232.

[7] B. Javadi, J. Abawajy, and R. Buyya, "Failure-aware resource provisioning for hybrid cloud infrastructure," *Journal of parallel and distributed computing*, 2012.

[8] R. Jhawar, V. Piuri, and M. Santambrogio, "Fault tolerance management in cloud computing: A system-level perspective," *Systems Journal, IEEE*, vol. 7, no. 2, pp. 288–297, 2013.

[9] D. R. Avresky, S. J. Geoghegan, and Y. Varoglu, "Evaluation of software-implemented fault-tolerance (sift) approach in gracefully degradable multi-computer systems," *Reliability, IEEE Transactions on*, vol. 55, no. 3, pp. 451–457, 2006.

[10] T. Alt, H. Appelshauser, S. Bablok, B. Becker, S. Chattopadhyay, C. Cheshkov, C. Cicalo, J. Cleymans, R. Fearick, H. Helstrup *et al.*, "Benchmarks and implementation of the alice high level trigger," *Nuclear Science, IEEE Transactions on*, vol. 53, no. 3, pp. 854–858, 2006.

[11] A. Imran, A. U. Gias, and K. Sakib, "An empirical investigation of cost-resource optimization for running real-life applications in open source cloud," in *High Performance Computing and Simulation (HPCS), 2012 International Conference on*. IEEE, 2012, pp. 718–723.

[12] M. Balazinska, H. Balakrishnan, S. R. Madden, and M. Stonebraker, "Fault-tolerance in the borealis distributed stream processing system," *ACM Transactions on Database Systems (TODS)*, vol. 33, no. 1, p. 3, 2008.

[13] J. A. Duraes and H. S. Madeira, "Emulation of software faults: a field data study and a practical approach," *Software Engineering, IEEE Transactions on*, vol. 32, no. 11, pp. 849–867, 2006.

[14] X. Liu, Q. Wang, S. Gopalakrishnan, W. He, L. Sha, H. Ding, and K. Lee, "Ortega: An efficient and flexible online fault tolerance architecture for real-time control systems," *Industrial Informatics, IEEE Transactions on*, vol. 4, no. 4, pp. 213–224, 2008.

[15] J. Dean and S. Ghemawat, "Mapreduce: simplified data processing on large clusters," *Communications of the ACM*, vol. 51, no. 1, pp. 107–113, 2008.

[16] Q. Zhang, L. Cheng, and R. Boutaba, "Cloud computing: state-of-the-art and research challenges," *Journal of Internet Services and Applications*, vol. 1, no. 1, pp. 7–18, 2010.

[17] J. H. Abawajy, "Fault-tolerant scheduling policy for grid computing systems," in *Parallel and Distributed Processing Symposium, 2004. Proceedings. 18th International*. IEEE, 2004, p. 238.

A Fast Recognition Scheme for Off-line Bangla Numerals

Md. Raihan ul Masood

Department of Computer Science and Engineering, Northern University Bangladesh
Holding no.13, Road no. 17, Banani, Dhaka-1213, Bangladesh
bdraihan@yahoo.com, bdraihan@nub.ac.bd

***Abstract*—The requirement of Bangla character recognition has become one of the prime attentions among the current researchers due to the increase of automated systems and usage of hand held devices. This paper presents a novel approach to recognize handwritten bangla numerals and addresses a robust feature extraction scheme that spawns 23- dimensional features based on the numeral's structure and topology. The recognition scheme is organized with a proposed decision tree that has been justified using entropy calculation. Considering inconsistency in individuals' writing style and the presence of significant curves and loops, the proposed feature extraction method restricts to less dimensional features of each numeral from 14600 data samples. The recognition time of this scheme is much lower than the existing procedures, since the preprocessing tasks have been performed during down sampling operation. The quick response time 13.04 ms and the higher accuracy (96.82%) explore a new era for the proposed scheme of being implemented in Bangla numeral LeadPad toys.**

Keywords* — *Character Recognition, Bangla Numeral, Down Sampling, Entropy, Feature Extraction

I. INTRODUCTION

Hand writing is one of the most effective ways of communication for general people. Though its popularity has reduced a little bit in this automatic Internet era, the efficient recognition scheme of handwritten scripts is desirable in every aspect of automated system. Bangla is the fifth-most popular language in the world. One set of bangle numerals from seven different handwritings are shown in Fig. 1. Automatic postal sorting, automatic bank cheque processing, children learning toys are some of the well known application areas for the hand written Bangla character recognition.

The history of character recognition is very rich and started around the mid twentieth century [1-2]. For the case of unconstrained handwritten numerals research has made impressive progress in Roman, Chinese and Arabic script in the recent time [3]. Bangla character recognition has also its well known literature for the last twenty years [4-6]. Pal and Chaudhuri worked with the recognition of unconstrained offline Bangla characters [7]. Islam et al. proposed grid method for recognizing On-line Bangla numerals by extracting different features which has an overall recognition rate of 96% [8]. Pal et al. used a novel scheme for Bangla handwritten numeral recognition which is mainly based on new features obtained from the concept of water overflow from the reservoir as well as topological and structural features of the numerals and it has overall accuracy of 92.8% from 12000 data [9]. Later on, highly efficient Bangla character recognition systems are proposed by artificial neural networks in [10-11] but they are computationally expensive. Y.Wen has presented a recognition system based on support vector machine, Kirsch mask and principal component analysis putting emphasis on response time [12]. They have extracted 1280-dimensional features from the input and then reduced to an unspecified number using principal component analysis (PCA).

This research takes into account on less computation in the preprocessing step employing a down sampling strategy and quick response time of recognition along with high accuracy. Though down sampling reduces the information, the off-line characteristics of the input numeral remain unchanged (as shown in fig. 2(a)).This research presents a novel decision tree for the fast recognition of the hand written Bangla numerals, considering the scope of implementing in interactive hand held devices. Since recognition scheme mostly depends on the feature extraction, all possible aspect of structural and topological characteristics of numerals have been considered and have found some unique features.

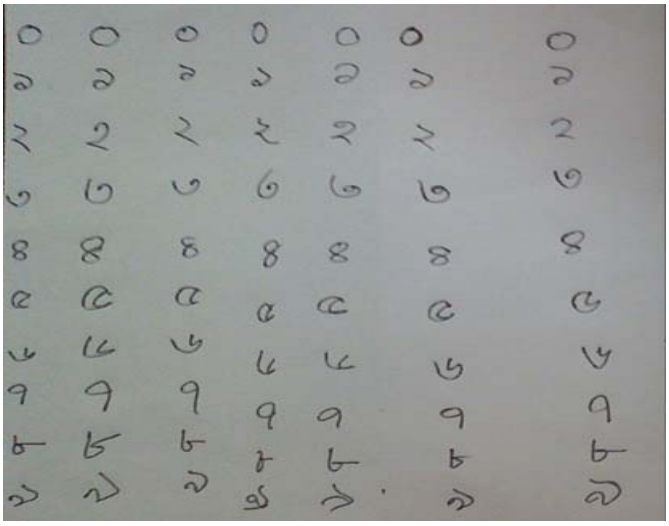


Fig. 1 Example of Bangla Numerals from 0 to 9 written by 7 different individuals.

The rest of this paper is organized as follows: Section 2 describes the steps of prepossessing technique needed for

recognition. Section 3 deals with detailed procedure of feature extraction. Section 4 comprises of proposed recognition scheme. Experimental results, achieved by the proposed method are shown in Section 5. Finally overall conclusion is drawn in section 6.

II. PREPROSSESSING

A. Data Acquisition

Acquisition of Bangla numerals are performed in two different ways: (i) with the help of mouse by direct drawing on the panel, and (ii) loading the image of the handwritten numeral. In order to get the numeral to be recognized, the user is given an interface (as shown in fig. 2). They can either draw the numeral in the white area of the writing panel or load an image of the numeral.

B. Warping the numeral with rectangle

Only single numeral input has been considered, hence segmentation of numeral has not been focused. An efficient segmentation technique has been mentioned through the concept of water reservoir [9]. From the input numeral four corner points are extracted, and rectangle is drawn from these four points (as shown in fig. 3 (a)).

C. Preparing the input image for feature extraction

In order to reduce the computational cost the wrapped input image is down sampled into 11x11 grids and features in different regions of interest are extracted(as shown in fig. 2).

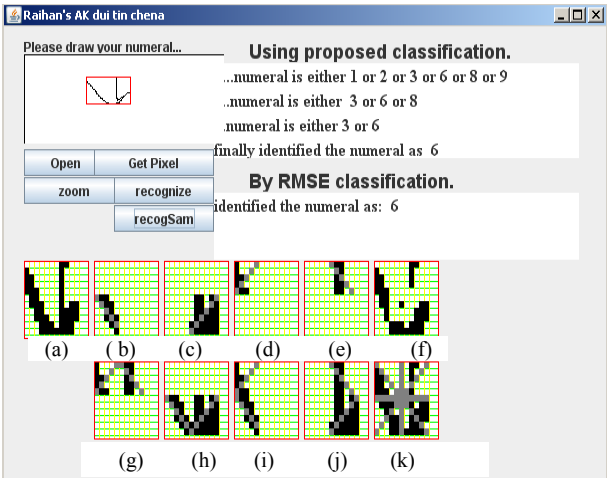


Fig. 2 User interface of the numeral recognition system

D. Pre-processing Computation

The pre-processing operation is performed as follows:

a. Compute the centre of gravity (later denoting as cg) of the positions of pixels in the bounded region considering the left- down point of the numeral as origin (as shown in fig. 3(b)).

b. Find number of pixels around the cg of one unit distance i.e. eight neighbours.

c. Find the occurrences of pixels of each section of the 8-half quadrant which is considered as the 8-symmetric segments of a circle (as shown in fig. 3(c)) and compute the sum-product of the occurrence of pixels in each row and column and their respective weight, EHQ_i is given by the expression:

$$EHQ_i = \sum_{i=0,j=0}^{i=3,j=3-i} n_{ij} . W_i \qquad (1)$$

where n_{ij} is represented as the occurrences of pixels in the jagged array (as shown in fig. 4(a) which is the 0th segment of the fig. 3(c)) and W_i is the weight matrix. The values of n_{ij} and the components of W are as follows

$$W = [15\ 4\ 1\ 1\ ], \quad n_{ij} = \begin{cases} 1, \text{ if pixel exists} \\ 0, \text{ otherwise} \end{cases}$$

The components of W have been chosen in such a way to give emphasis on the weight of the pixels near the circumference. The justification of the chosen value of W is existence of one pixel in the 2nd row produce larger value (EHQ= $n_{1,j}$ W=1×4=4) than all the three pixels in the 3rd and 4th row (EHQ= 1×1+1×1+1×1=3), again the existence of one pixel in the 1st row (EHQ= $n_{0,j}$ W=1×15=15) will produce the same as produced by the six pixels in the 2nd, 3rd and 4th row (EHQ=1×4+1×4+1×4+1×1+1×1+1×1=15) (fig. 3(c) and fig.4(a)).

d. Compute the number of pixels inside and outside the diagonal of the four quadrants (as shown in fig. 5(a)).

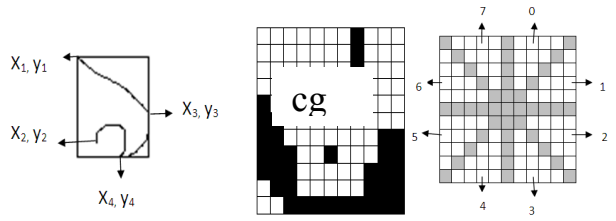


Fig. 3 (a) Rectangle boundaries around the numerals (b) Showing the cg of numeral 6 (c) Eight region of interest for EHQ

e. Find the discontinuity or cavity (i.e. absence of pixel in between pixels in the same row or column).

W_0 = 15	W_0 = 15	W_0= 15	W_0 = 15
W_1 = 4	W_1 = 4	W_1 = 4	
W_2 = 1	W_2= 1		
W_3 = 1			

EHQ_0 = 0×15 + 1×15 + 1×15 + 0×15 + 0×4 + 1×4 + 0×4 + 0×1 + 1×1 + 0×1 = 35, using(1).

Fig. 4 (a) Jagged array showing the different values of W (enlarged picture the 0th portion of Fig.3(c)) (b) Sample EHQ calculation (The encircled portion of of Fig. 5(c))

III. FEATURE EXTRACTION

The features of each numerals have been extracted depending on its evidence by 14600(=146x10x10) data samples. In this proposed scheme eight major areas have been taken into consideration for effective feature extraction. After testing the data samples the average values of $EHQ_{i=0...7}$ for all numerals have been taken (as shown in fig.6). From the average values of EHQ_i, distinctive classes of numeral have been formed such as the EHQ_7 of numeral 3 and 6 are significantly low from the other numerals. For mathematical interpretation of the term 'low' has been used as a threshold value $T_1 <= 15$, which is the largest value of W in (1).

Along with some fixed topological and structural property viz. numeral 5 always has concave property in the right-half portion, some other crucial features have been extracted. The computation described in section IIDe has revealed the discontinuity feature that numeral 1 does not have discontinuity in either region 1 or region 2 (fig. 5(b) and 6(c)) whereas numeral 9 always has (fig. 6(a) and 6(b). This discontinuity feature can easily separate the closely related numeral pair such as numeral 1 and numeral 9. The convex property in eight different regions (fig. 2(b, c, d, e, g, h, i, j)) and the discontinuity in another six areas of interest (fig. 5(b)) have also taken into account for effective recognition. Therefore, along with EHQ in eight different region and cg, twenty three dimensional features of each numeral have been extracted.

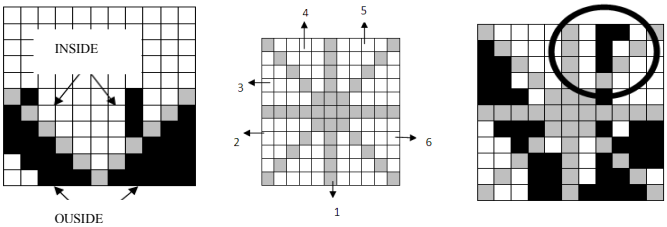


Fig. 5 (a) Showing lower half inside concavity of numeral 6 (b) Six area of interest to find discontinuity (c) Numeral 6 for EHQ calculation

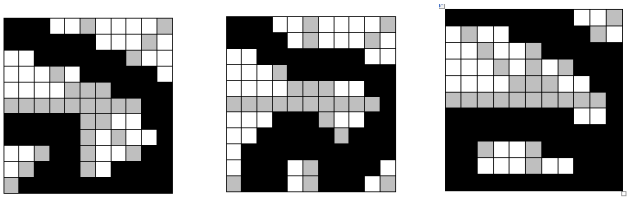


Fig. 6 (a) Numeral 9 having discontinuity in region 2 (Fig. 5(b)) (c) Numeral 9 having discontinuity in region 1 (Fig. 5(b)) (c) Numeral 1 has no discontinuity in either region 1 or 2 (Fig. 5(b))

	EHQ0	EHQ1	EHQ2	EHQ3	EHQ4	EHQ5	EHQ6	EHQ7	cg
১	38	48	66	66	51	27	6	61	2
২	25	21	12	55	29	31	9	55	2
৩	34	56	64	64	31	23	63	8	3
৪	54	43	57	59	55	34	57	69	6
৫	48	24	45	66	52	53	63	58	2
৬	30	19	62	64	33	24	57	8	3
৭	52	52	54	18	2	7	67	65	3
৮	5	29	12	26	62	63	64	29	4
৯	31	44	64	52	26	28	14	53	4
০	60	59	61	57	58	62	59	56	1

Fig. 6 the values of EHQ and cg of 10 numerals from 14600 data

IV. NUMERAL RECOGNITION

Two major aspects, accuracy and response time must be considered in recognition process. Considering this fact the proposed scheme uses two types of recognition processes using the extracted features.

a. Root Mean Square Error (RMSE): This classifier first find the root mean squared error by the following equation:

$$e_i = \sqrt{\frac{1}{n}\sum_{j=0}^{n}(S_{ij} - T_j)^2} \quad (2)$$

Where, $S_{i=0...9,\ j\ =\ 0....23}$ represents the features of printed 10 numerals and $T_{j=0...23}$ is the features of the test numeral. Then the classifier computes the minimum of e_i and decides that the tested numeral belongs to i^{th} class.

b. A Decision Tree: This is the proposed classifier for numeral recognition based on a decision tree (as shown in fig.7). The decision tree has also been justified using entropy calculation using the following equation:

$$\text{Entropy} = \sum p_i \log_2 p_i \quad (3)$$

where i=1, 2, 3,……,n and p is the probability of chance.

For constructing the decision tree some notations are discussed as follows:

- $EHQ_{i=0\ to\ 7}$, T_1 and cg have already been defined.
- $IsCont_{i=0\ to\ 5}$, presents a boolean function. It returns true if continuity exists in any of six region of interest (as shown in fig. 5(b)) else returns false.
- T_2, represents a boolean value. It returns true if the number of pixels present in the eight neighbouring position of centre of gravity is more than or equal to 4 else return false.

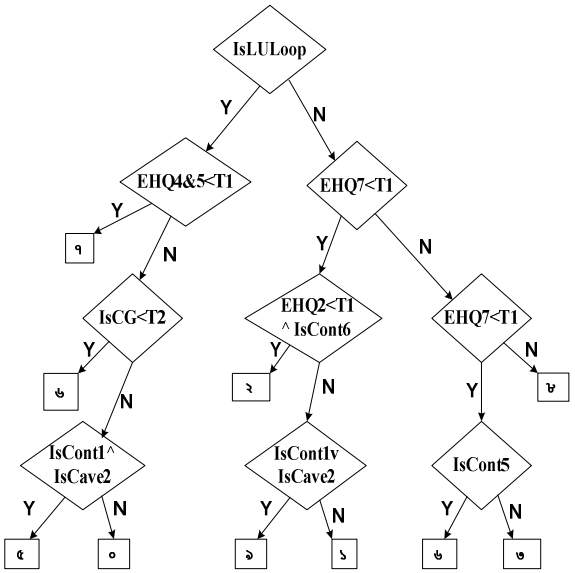


Fig. 7 Decision tree for the recognition process of bangle numerals.

- $IsCave_{i=0\ to\ 3}$(0-left, 1-top, 2-right and 3-bottom), presents a boolean function. It returns true if number of pixel inside is less than the number of pixel outside (as shown in Fig. 5(a)) else returns false.

- IsLULoop presents another boolean function that returns true if both $IsCave_0$ and $IsCont_3$ are both true else returns false.

V. Result and Discussion

The proposed scheme has been tested on 14600 numerals from 146 different individuals. Each individual has tried 10 different writing styles for each numeral. They used mouse to draw numeral through the interface (as shown in fig. 2). It is noted that the overall accuracy of the recognition scheme is 96.82% (as shown in table I).

TABEL I

Numeral	No. of sample data	Average response time(ms)	Accuracy
0	1460	11.4	98.50%
1	1460	14.3	97.71%
2	1460	13.6	98.50%
3	1460	15.3	94.20%
4	1460	15.3	93.70%
5	1460	12.7	98.50%
6	1460	11.8	95.70%
7	1460	12.2	99.20%
8	1460	12.6	95.80%
9	1460	11.2	96.40%
	Overall	13.04	96.82%

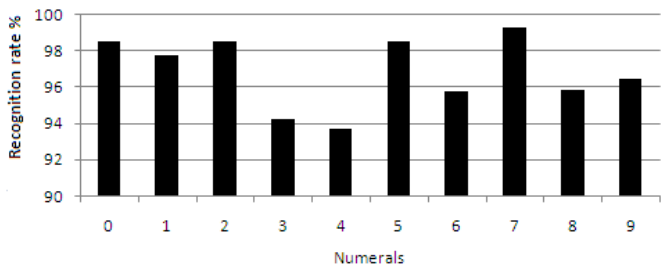


Fig. 8 Chart of success rate of recognition of bangla numerals.

The experiment shows that the highest accuracy achieved for numeral 7(as shown in table I). It is because its unique lower half features. The failed rate of 0.8% happened owing to the vulnerability of the proposed scheme to unnecessary segments (as shown in fig. 9(a)). It can be solved using pruning technique.

Low rate of accuracy of numeral 4 has occurred because it's large deviation of orientation and large discontinuity in the upper loop (as shown in fig. 9(c)).

The proposed scheme has achieved high accuracy for the most confusing pairs (numeral 1 and 2, numeral 1 and 9) [5] [1] by introducing the discontinuity features (section IIDe). This feature is also responsible for the high accuracy of numeral 2, since extra looping (as shown in fig. 9(b)) has been taken care of by this feature.

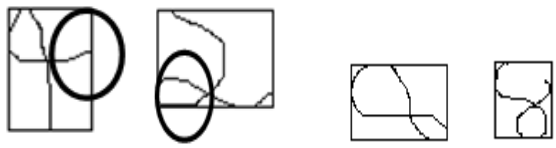


Fig. 9 (a) Numeral 7 with large unwanted segment (b) Numeral 2 with extra loop like numeral 1 (c) Numeral 4 with flatted orientation and large discontinuity in the top region

Though different researchers use different data sets from their respective space and time, main objective is to achieve the high accuracy of recognition numerals. Comparison among some methods of recognition of Bangla numerals is shown in fig. 10.

The response time includes down sampling the numeral image, feature extraction and the recognition phase. Since the proposed method mainly focuses on a novel Bangla numeral recognition approach, it only deals with a single numeral. Image segmentation of over lapping or touched numerals is not done in this experiment as it is considered in [9][14]. Moreover, after feature extraction all the decision making process towards the numeral recognition is done in unit time, hence a quick response time is achieved. Though the base-line of the comparison is not same, in terms of response time of Bangla numeral recognition a comparison is exposed in fig. 11.

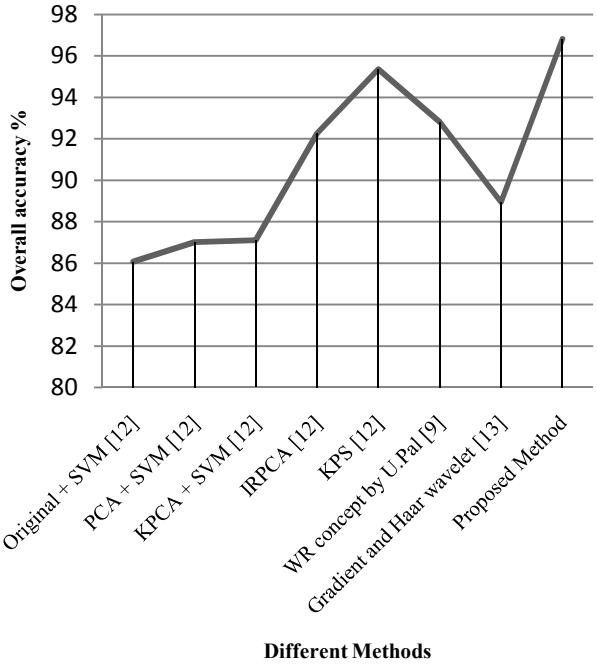


Fig. 10 Numeral recognition accuracy of different methods

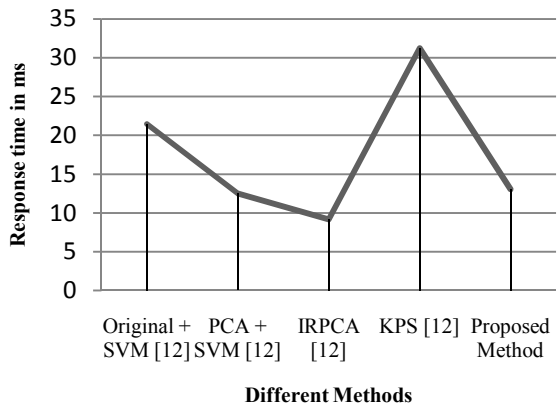


Fig. 11 Recognition response time of different methods in ms.

The experiment is done on the system with Pentium dual core processor of 1.8 GHz and 1GB ram and the implementation is done using JDK 1.6.

VI. CONCLUSIONS

This paper presents a novel approach for feature extraction of off-line Bangla numerals. The feature is extracted on the basis of pixels distribution in the eight equivalent segments of the rectangle wrapped around the test numerals. Considering the discontinuity and the curvature properties of the numerals, the variation in orientation of the handwritten numerals has been taken care of. Based on the features, the proposed scheme uses two classifiers for recognition process: (i) Root Mean Square Error and (ii) A decision tree. Sufficient test samples reveal the accuracy of this new approach. At present this scheme only deals with faster recognition. It will continue its journey by proposing a new recognition scheme of on-line bangla numeral as well as bangla character with more accuracy.

REFERENCES

[1] L. D. Earnest.Machine reading of cursive script.In IFIP Congress, pp 462-466, Amsterdam, 1963.

[2] M. Eden. Handwriting and pattern recognition. IRE Trans. Information Theory, 8, 1962.

[3] R. Plamondon and S. N Srihari. On-line and off-line handwritten recognition: A comprehensive survey. IEEE Trans. on PAMI, 22: 62-84, 2000.

[4] M. A. Sattar and S. M. Rahman. An experimental investigation on Bangla character recognition system. Bangladesh Computer Society Journal, Vol 4. pp. 1-4, December 1989.

[5] M. A. Sattar. Recognition of printed Bengali character by decision theoretic method of pattern recognition. Master's thesis, Dept. of CSE, BUET, 1990.

[6] M. A. K. Mia. Recognition of printed Bangla characters by syntactic method of pattern recognition. Master's thesis, Dept. of CSE, BUET, 1992.

[7] U. Pal and B. B. Chaudhuri. Automatic recognition of unconstrained offline Bangla hand-written numerals. Advances in multimodal Interfaces, (LNCS-1948), pp. 371-378, 2000.

[8] M. M. Islam, S. Rahman, A. N. M. E. Rafiq and M.M. Islam, On-line handwritten Bangla numeral recognition by grid method, In Proceedings of 2nd ICECE, 2002.

[9] U. Pal, A. Belaid and B. B. Chudhuri. A system for Bangla handwritten numeral recognition. Available: www.loria.fr/~abelaid/publis/kbcs-cam-final.pdf.

[10] M. R. Khan, S. M. A. Rahman and M. M. Alam. Bangla handwritten digits recognition using evolutionary artificial neural networks. 7th International Conference on Computer and Information Technology, (ICCIT), 2004.

[11] D. Das and R. Yasmin. Segmentation and recognition of unconstrained Bangla handwritten numeral. Asian Journal of Information Technology. Vol. 5(2), pp. 155-159, 2006.

[12] Ying Wen,Yue Lu,Pengfei Shi. Bangla numeral recognition system and its application to postal automation. Pattern Recognition 40 (2007) 99 – 107

[13] Sekhar Mandal, Sanjib Sur, Avishek Dan, Partha Bhowmik. Handwritten Bangla character recognition in machine-printed forms using gradient information and Haar wavelet. International Conference on Image Information Processing(ICIIP-2011).

[14] Md. Aktaruzzaman, Md. Farukuzzaman Khan, Dr. Ahsan – ul – Ambia. A new technique for segmentation of handwritten numerals string of Bangla language. I.J. Information Technology and Computer Science, 2013,05,38-43.

Study of Different Forecasting Models On Google Cluster Trace

Md. Rasheduzzaman, Md. Amirul Islam, Tasvirul Islam,
Tahmid Hossain, Rashedur M Rahman

Department of Electrical Engineering and Computer Science,
North South University,
Plot-15, Block-B, Bashundhara, Dhaka 1229, Bangladesh
{rashed091nsu, amirul507, aurgho350, tahmid}@gmail.com, rashedur@northsouth.edu

***Abstract*—Workload prediction in cloud system is an important task and it helps in efficient resource allocation by minimizing cost and thus maximizing the profit. In this paper we analyze a large scale production workload trace (version 2) which is recently made publicly available by Google. The main objective of our research is to design and compare different forecasting models. We develop models through Adaptive Neuro-Fuzzy Inference System (ANFIS), Non-linear Autoregressive Network with Exogenous inputs (NARX), and Autoregressive Integrated Moving Average (ARIMA). Finally, we compare these three prediction models to find out the best one. Performance of forecasting techniques is measured by two popular statistical metrics, i.e., Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE). The experimental result demonstrates that NARX model outperforms other models, e.g., ANFIS and ARIMA.**

***Keywords*—Workload; Google Cluster Trace; Neural Network; ANFIS; ARIMA.**

I. Introduction

Cloud computing is a way to increase capacity or add capabilities without investing in new infrastructure. Recently, cloud computing has become more popular to the different industries and company. System log data produced by different industries are increasing rapidly. Each data center always tries to minimize their resource cost and maximize their profits. If we can predict the amount of resources needed in the future and accordingly make an allocation, this could help us to increase the profit and maximize performance. Time series analysis can be used as a forecasting method. It is heavily used in the field of economics and other decision making systems. Time series analysis could be a nice tool to estimate the future workload or resource usage. Our study shows an approach to build a prediction model using ANFIS, NARX, and ARIMA. Specifically, our goal is to find a better predictive model for Google cluster data. In this paper we tried to find the forecasting model that works the best for real world cluster workload data.

The rest of the paper is organized as follows. Section II discusses about related work on ANFIS, NARX, and ARIMA model. The systematic description of the ANFIS, NARX, and ARIMA model is given in section III. Section IV discusses the experiments and presents a detailed comparison of experimental results of these three models. Finally, we draw the conclusion from our experiments and give direction of future research in section V.

II. Related work

A lot of researches have been conducted in forecasting time series. These studies have provided a better understanding of what could happen in the future. Wang, Chau, Cheng, and Qiu, [18] compared different Artificial Intelligence (AI) (such as SVM, ARIMA, ANFIS, ANN and GP) methods in order to build effective hydro-power resource management and scheduling. The authors used four different statistical performance measurement techniques to validate different AI methods. They showed that AI methods are more powerful compared to traditional time series forecasting model. They found that SVM has better result in both training and validation stage. They also report that ANFIS shows different performance in training and validation stage. GP has better result in validation phase. Contreras, Espínola, Nogales, and Conejo [6] carefully studied the performance impact of ARIMA model on next-day electricity price forecasting. The authors proposed two different ARIMA models for predicting electricity price in Spain and California. In their research, they showed that the ARIMA model for Spain required only 5 hours to forecast future price, where California required only 2 hours. Daz-Robles et al. [8] designed a hybrid model by combining ARIMA and ANN model for forecasting air quality in urban areas. They showed that traditional Box-Jenkins time series (ARIMA) model and multi-linear regression (MLR) models have limited accuracy. Khashei and Bijari [12] built a novel predictive model by combing ARIMA and ANN model which gave better accuracy. They tested their model with three real-data set and argued that in case of higher performance, this model could be used as an appropriate alternative. Boyacioglu, and Avci [3] examined the performance of ANFIS model for stock market return in order to determine whether ANFIS model is capable of giving better accuracy. They tested their ANFIS model on Istanbul Stock Exchange (ISE). Then they showed that ANFIS was able to predict 100 indices of ISE with 98.3% accuracy.

III. MODELS USED

A. The ANFIS model

Adaptive Neuro-Fuzzy Inference System (ANFIS) is one kind of neural network which is based on fuzzy inference system. Its inference system corresponds to three type components. These are: (I) set of fuzzy IF-THEN rules that have learning capacity (II) database for defining membership function and (III) an inference system. ANFIS uses two learning algorithms (1) back-propagation and (2) least square estimation to learn the parameters. The ANFIS architecture is shown in Fig. 1. The circular nodes represent nodes that are fixed whereas the square nodes are nodes that have parameters to be learnt. A Two Rule Sugeno ANFIS has rules of the form

$$\text{if } x \text{ is } A_1 \textbf{ AND } y \text{ is } B_1 \text{ then } f_1 = p_1x + q_1y + r_1 \quad (1)$$

$$\text{if } x \text{ is } A_2 \textbf{ AND } y \text{ is } B_2 \text{ then } f_2 = p_2x + q_2y + r_2 \quad (2)$$

Here, x and y are crisp inputs to the node. A_i and B_i are the linguistic labels. The linguistic variables are characterized by membership function p_i and q_i, where x_i and y_i are crisp inputs to the node A_i and B_i. The linguistic variables are characterized by membership function p_i and q_i. For the training of the network, there is a forward pass and a backward pass. The forward pass propagates the input vector through the network layer by layer. In the backward pass, the error is sent back through the network in a similar manner to back propagation. The ANFIS structure consists of five layers. The description of each layer is given below. More details could be found elsewhere [3, 6, 15, 17, 19].

1) Layer 1:

Each node in this layer generates the membership function of a linguistic label. The ith node may perform the following operation: where x is the input of ith node, where a_i, b_i, c_i are parameters to be learnt. These variables are used to adjust the shape of membership functions.

$$O^1_i = \mu_A(x) = \frac{1}{1 + [(\frac{x-c_i}{a_i})^2]^{bi}} \quad (3)$$

2) Layer 2:

The firing strength of each rule is calculated by:

$$O^2_i = w_i = \mu_{Ai}(x) \times \mu_{Bi}(y),\ i = 1, 2 \quad (4)$$

3) Layer 3:

Layer 3 calculates the ratio of ith rule's firing strength to the sum of all rule's firing strength:

$$O^3_i = \hat{w}_i = \frac{w_i}{w_1 + w_2},\ i = 1, 2 \quad (5)$$

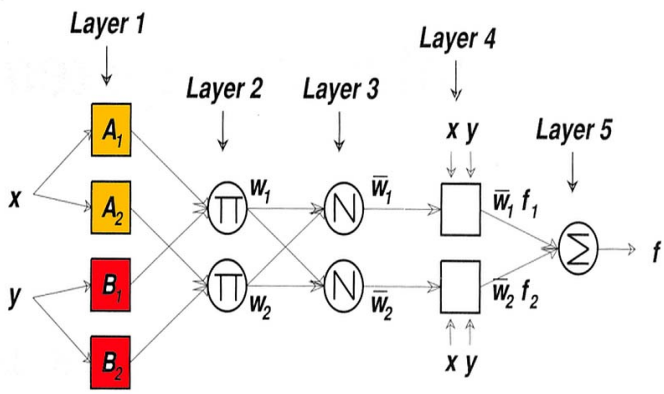


Fig. 1. ANFIS Architecture

4) Layer 4:

Every node in this layer has the following function:

$$O^4_i = \hat{w}_i f_i = \hat{w}_i\,(p_ix + q_iy + r_i) \quad (6)$$

5) Layer 5:

Layer 5 aggregates the overall output as the summation of all incoming signals:

$$O^5_i = \sum \hat{w}_i f_i = \frac{\sum_i \overline{w_i} f_i}{\sum_i w_i} \quad (7)$$

B. The NARX model

Recurrent Neural Networks (RNN) are used for modeling both powerful nonlinear or linear dynamic systems. A widely used example of the recurrent dynamic network is the non-linear autoregressive network with exogenous inputs (NARX). The NARX model is based on the linear ARX model, which is commonly used in time-series modeling. Typically when a non-linear regression vector is selected as for ARX models, the model structure is called NARX. Fig. 2 shows a typical NARX model architecture. More details of this model could be found elsewhere [1, 2, 5, 10].

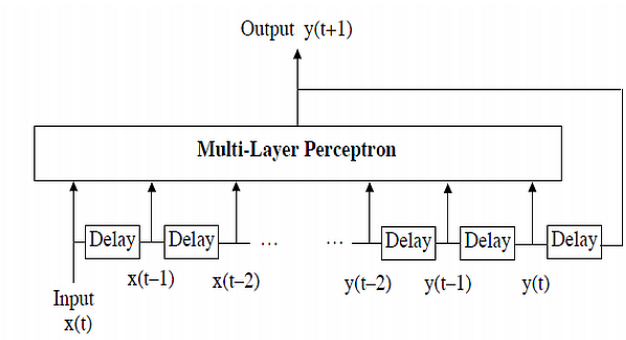


Fig. 2. NARX Architecture

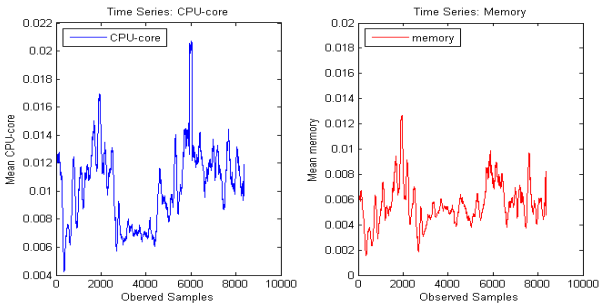


Fig. 3. Original Data Plot

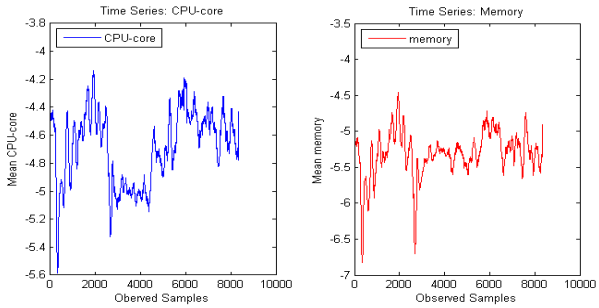


Fig. .4. Log Data Plot

The input to an NARX model consists of past output data, and past input data. The NARX model is represented by the nonlinear difference equation of the form:

$$y(t) = F[x(t), \ldots, x(t-d), y(t-1), \ldots, y(t-d)] \quad (8)$$

Where, F is the nonlinear part of the equation, x (t) and y (t) represent the input and output of the model at time t. This network also uses tapped delay lines (d) to store previous values of the input, x(t) and output, y(t) sequence. When the nonlinear function is modeled with a Multilayer Perceptron (MLP), the resulting network is called a NARX recurrent neural network. A NARX model becomes a Time Delay Neural Network (TDNN), when the past output lag becomes zero. The TDNN model is represented by the equation of the form:

$$y(t) = F[x(t), x(t-1), \ldots, x(t-d)] \quad (9)$$

C. The ARIMA model

The term ARIMA stands for "Autoregressive Integrated Moving Average" (also known as Box-Jenkins methodology). ARIMA is widely used for time series forecasting. ARIMA can be used to represent both stationary as well as non-stationary time series. An ARIMA model consists of three different components. The simplest form of Box-Jenkins model is called "Autoregressive"' (AR), which is similar to linear regression. A pth order autoregressive (AR) model has the form:

$$z_t = C + \varphi_1 z_{t-1} + \varphi_2 z_{t-2} + \ldots + \varphi_p z_{t-p} + a_t \quad (10)$$

Where C is the constant level, $z_{t-1} + z_{t-2} + \ldots + z_{t-p}$ are past series values (lags), the φ_s are coefficients to be estimated and a_t is a random variable with mean zero and constant variance. The second type of Box-Jenkins model is called a "Moving Average" (MA) model. Both AR and MA model have similar forms of the equation, but their purposes are different. A qth order moving average (MA) model has the form:

$$z_t = -\theta_1 z_{t-1} - \theta^2 z_{t-2} - \theta^3 z_{t-3} - \ldots + a_t \quad (11)$$

Where θ_s are coefficients to be estimated and a_t is an error term. The negative term in front of the parameter is used for convention only. The combined model of AR and MA is denoted by ARMA(p,q), which can be used to forecast different variety of stationary time series. Most real series are not stationary. In order to make the series stationary we used a differencing method to remove the trends and cycles of the series. These three together become the ARIMA(p,d,q) model. Here d stands for differencing order. ARIMA model heavily relies on autocorrelation (ACF) and partial autocorrelation function (PACF). ACF and PACF is used to determine the p orders and q orders of difference series (or stationary series). More details of this model could be found elsewhere [4, 6, 8, 11, 13].

IV. Experimental Study

A. Dataset description and preparation

Google has recently made publicly available 2 traces of their compute clusters [9]. Those traces describe the workload of their compute cell. We use trace version 2 which consists of 29 days worth of cell information of a single cluster. Before we apply different forecasting models on Google compute cluster, workload and resource usage (such as CPU-core, memory, etc.) of jobs should be understood properly. It is very important that we set a preliminary model for our analysis. Therefore, we begin with preparing the dataset. In our experiment we only consider the task usage table. We took mean of CPU-core and memory usage of every job seen in five minute interval for 29 days. From each day we have 288 samples and totally we have 8352 sample data point.

Fig. 3 shows the original data set that we prepared for our experiment. For better view we plotted the data in smoothing scale using Loses smoothing with a span. In order to reduce the mean and variance of the series we apply data transformation techniques. We call these transformation variances-stabilizing because their purpose is to make variances the same. We found that in our original data the standard deviation is proportional to mean, because of that we apply the natural logarithmic transformation. Fig. 4 shows the new logarithmic transformed data series.

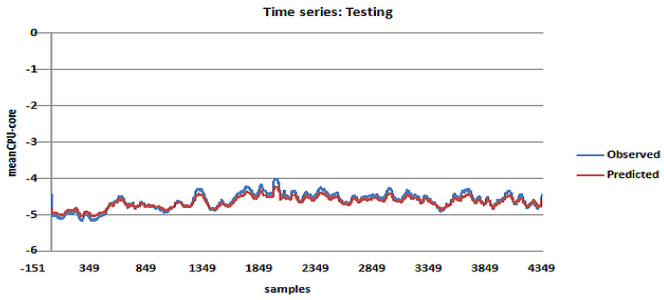


Fig. 4. ANFIS CPU-core Testing Performance

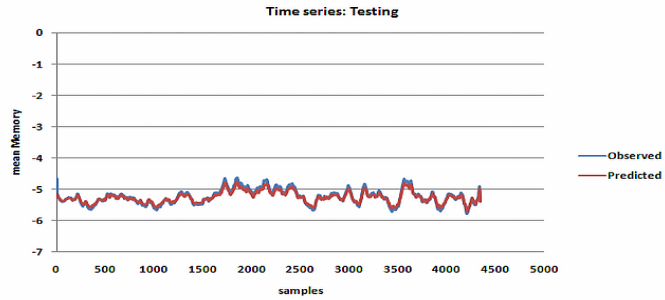


Fig. 5. ANFIS Memory Testing Performance

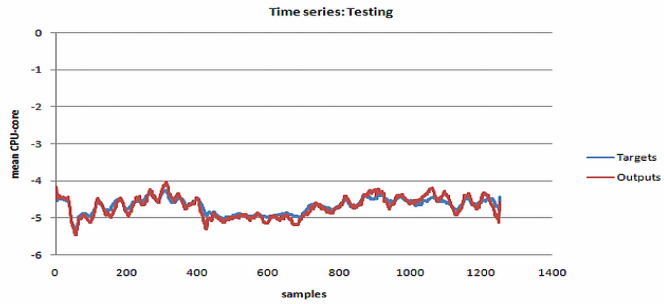


Fig. 6. NARX CPU-core Testing Performance

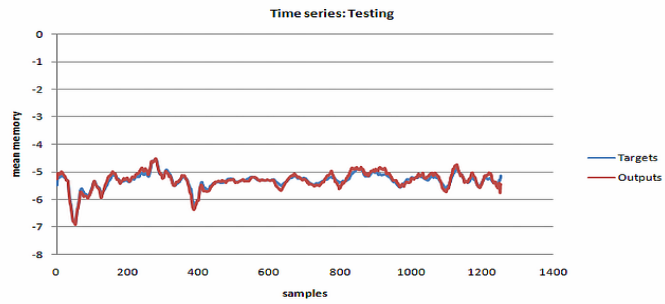


Fig. 7. NARX Memory Testing Performance

B. Performance criterion

The forecasting performance of the ANFIS, NARX, and ARIMA model is measured with two statistical methods, such as the Root Mean Squared Error (RMSE) and the Mean Absolute Error (MAE), which are widely used for performance measurement metric for time series forecasting.

$$\text{MAE}=\frac{1}{N}\sum_{t=1}^{n}\left|\frac{y_t-\overline{y_t}}{y_t}\right| \tag{12}$$

$$\text{RMSE}=\sqrt{\left(\frac{1}{N}\sum_{t=1}^{n}\left(y_t-\overline{y_t}\right)^2\right)} \tag{13}$$

Where N is the number of samples in the data set, y_t is original observation and $\overline{y_t}$ is the predicted value. The best model has MAE and RMSE error rate relatively small.

C. Results

1) ANFIS Experimental studies

The first stage of the experiment is to select the input parameters and prepare the dataset for ANFIS input layer. In our experiment we only used CPU-core usage and memory usage of the resource usage table from Google Cluster dataset. ANFIS uses a hybrid learning algorithm combines with back-propagation gradient descent and a least-square method to identify the parameters needed to reduce the network error. We prepared two different input datasets for ANFIS: (1) Two-input and (2) Three-input. We also used two and three membership functions for each input set. The membership functions are

TABLE I: RESULT OF DIFFERENT ANFIS STRUCTURE

Member-ship Function	Memb-ership Number	Perfor-mance Criteri on	Input (CPU core)		Input (Memory)	
			2	3	2	3
Gaussian	2	MAE	*.0713*	*.0685*	*.0586*	*.0577*
		RMSE	*.4133*	*.3989*	*.3889*	*.3827*
	3	MAE	.0722	.0729	.0585	.0582
		RMSE	.4316	.5044	.3886	.3881
Generalized Bell	2	MAE	.0709	**.0685**	.0586	**.0577**
		RMSE	.4100	**.3988**	.3888	**.3824**
	3	MAE	.0714	.0728	.0586	.0582
		RMSE	.4193	.4648	.3890	.3865

Generalized Bell and Gaussian. We divided each input set into two series. The first series was used for training the network and the other series was used to test the ANFIS network. For training we used 4000 samples and for testing we used 4352 samples. The ANFIS network performance is measured with RMSE and MAE.

Matlab scripting programming language was used to develop our Neuro fuzzy model. A training data set that contains the desired input output pairs of the system were presented to ANFIS toolbox for training. After finishing applying different configuration on ANFIS network, we found that two generalized bell membership functions for three input set has the lowest MAE and RMSE for both CPU-core and memory usage dataset. Fig. 5 and Fig. 6 show the best testing performance of CPU-core and memory usage. Table I summarizes the performance of two membership functions. The lowest error is highlighted in bold letter in the Table I.

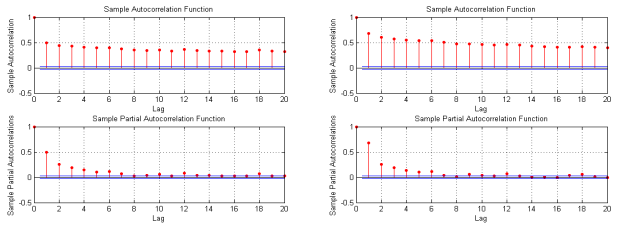


Fig. 8. ACF and PACF Plot of Logarithmic Time Series of CPU and Memory Usage

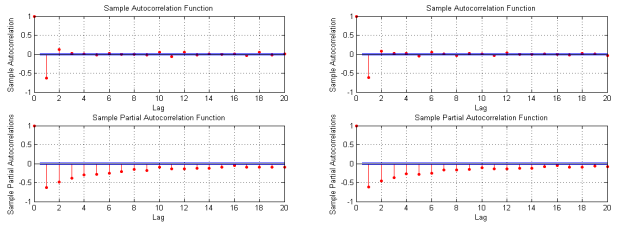


Fig. 9. ACF and PACF Plot of 2nd Order Differencing Series of CPU and Memory Usage

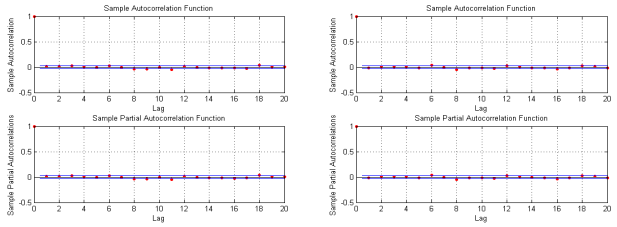


Fig. 10. ACF and PACF Plot of Residuals of Best ARIMA Model of CPU and Memory Usage

2) NARX experimental studies

In our experiment we only used the Levenberg-Marquardt algorithm for training. We used 75% of our data for training and 15% for validation and another 15% for testing. There is no commonly accepted method for selecting the number of hidden layers or the number of input delay for NARX model. We used a different structure of NARX model to find out the best structure. Fig. 7 and Fig. 8 show testing result of best NARX model. Table II and Table III summarize the result of the different NARX model structure applied to CPU-core and memory usage.

After finishing the training, validation, and testing of NARX model, we pick the best NARX model based on the lowest testing error rate. We found that NARX model has lowest MAE and RMSE error rate with delays 1:2 with hidden neuron 9 and 7 respectively for CPU-core and memory.

3) ARIMA experimental studies

For the ARIMA model we used 4000 samples as training and rest as test samples. For these data which ARIMA model is appropriate for the time series is unknown. But usually the value of p, d, q over 3 are rarely needed. Often time different ARIMA model gives the same forecasting result. Therefore, there is flexibility in our choice. In order to select the best ARIMA model we plot the data. Fig. 4 shows the plot of logarithmic Google cluster time series for CPU-core and memory. We took the natural logarithm of the original time series in order to reduce the mean and variance of the series.

To see whether the series is stationary we plot ACF and PACF of logarithmic time series. From the Fig. 9 we see that both ACF and PACF are declining very slowly. It implies that input series is not stationary. Also a large number of ACF and PACF indicate that series is non-stationary. In order to make it stationary we need differencing. Stationary series implied that either ACF of PACF drops to zeros very quickly. After applying the second order differencing on input logarithmic series, the series becomes stationary series. If the lags of ACF die off quickly, it implies that the series follows a low-order AR process. If the lags of ACF drop to zero after a few lags, it implies that the series follows a low-order moving-average process. From the Fig. 10 it is quite difficult to find appropriate AR(p) or MA(q) value just by looking at ACF and PACF plot. Therefore, we apply minimizing information criterion such as (AIC or BIC). Result of different ARIMA model is given in Table IV. Fig. 11 shows the best fitted ARIMA model residuals plot of ACF and PACF.

TABLE II: NARX TESTING RESULT OF CPU-CORE

Delays	Error Method	Number of Hidden Neurons			
		7	8	9	10
1:2	MAE	.0667	.0636	**.0618**	.0646
	RMSE	.3900	.3760	**.3700**	.3828
1:3	MAE	.0645	.0658	.0622	.0628
	RMSE	.3883	.3871	.3848	.3768

TABLE III: NARX TESTING RESULT OF MEMORY

Delays	Error Method	Number of Hidden Neurons			
		7	8	9	10
1:2	MAE	**.0519**	.0532	.0547	.0553
	RMSE	**.3456**	.3548	.3682	.3690
1:3	MAE	.0537	.0529	.0519	.0527
	RMSE	.3602	.3511	.3498	.3565

TABLE IV: ARIMA TESTING RESULT

ARIMA Model	Performance Criterion	CPU-core	Memory
ARIMA(1,2,2)	MAE RMSE	.1821 .9612	28.2838 192.1374
ARIMA(2,2,2)	MAE RMSE	.1656 .8981	24.1913 163.4591
ARIMA(2,2,0)	MAE RMSE	50.3027 261.2828	13.4134 103.3555
ARIMA(2,0,1)	MAE RMSE	**.0903** **.4871**	**.0831** **.5105**
ARIMA(1,0,1)	MAE RMSE	.0911 .4904	.0837 .5143
ARIMA(2,1,1)	MAE RMSE	.0997 .5925	.1359 .8661
ARIMA(1,1,1)	MAE RMSE	.1026 .6061	.1361 .8672
ARIMA(1,2,1)	MAE RMSE	3.2688 18.6161	21.7256 147.9280

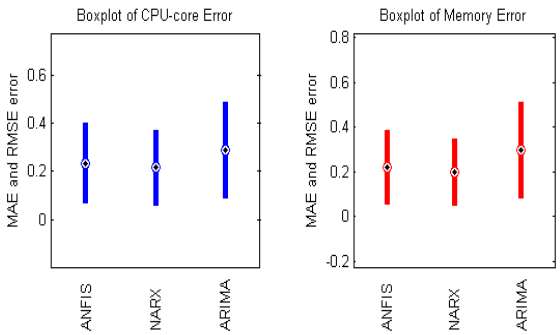


Fig. 12. Performance Comparison

TABLE V: BEST RESULT FOR CPU-CORE USAGE

Criterion	ANFIS	NARX	ARIMA
MAE	0.0685	0.0618	0.0903
RMSE	0.3988	0.3700	0.4871

TABLE VI: BEST RESULT FOR MEMORY USAGE

Criterion	ANFIS	NARX	ARIMA
MAE	0.0577	0.0519	0.0831
RMSE	0.3824	0.3700	0.4871

D. Comparison of performances

After finishing analyzing three different forecasting methods we compare the performance of ANFIS, NARX, and ARIMA model. The best test result of ANFIS is compared with NARX and ARIMA model in order to get the best forecasting model of Google compute cluster. The performances of the best models developed by ANFIS, NARX, and ARIMA models for forecasting data sets are summarized in Table V and VI.

Table V and Table VI summarize the best result achieved by ANFIS, NARX, and ARIMA model with their corresponding MAE and RMSE error rate for Google compute cluster data. Fig. 12 shows the box-plot of error rate of three different models for both CPU-core and memory usage. From the box-plot it can be shown that NARX model has the lowest error rate compared to ANFIS and ARIMA.

V. CONCLUSION

In this paper we present different forecasting models to predict the workload of Google's compute cluster time series. The forecasting models that we design and analyze include the ANFIS, NARX, and ARIMA method. Our analysis for workload prediction shows that NARX model has low training and testing error. The ANFIS model gives almost same performance compared to NARX, but ARIMA has poor training and testing error compared to ANFIS and NARX model.

REFERENCES

[1] P. Amani, M. Kihl, and A. Robertsson, "Narx-based multi-step ahead response time prediction for database servers", In the Proceedings of 11th International Conference on Intelligent Systems Design and Applications(ISDA), pp. 813 – 818, Spain, 2011.

[2] S. H. Arbain and A. Wibowo, "Neural networks based nonlinear time series regression for water level forecasting of Dungun river", Journal of Computer Science 8(9):1506 - 1513, 2012.

[3] M. A. Boyacioglu and D. Avci, "An adaptive network-based fuzzy inference system (anfis) for the prediction of stock market return: The case of the Istanbul stock exchange", Expert System with Applications, 37(12):7908 - 7912, 2010.

[4] C. Christodoulos, C. Michalakelis, and D. Varoutas, "Forecasting with limited data: Combining ARMA and diffusion models", Technological Forecasting and Social Change, 77(4):558 - 565, 2010.

[5] J. T. Connor, R. D. Martin, and L. E. Atlas, "Recurrent Neural Networks and Robust Time Series Prediction", IEEE Transactions on Neural Networks, 5(2):240 - 254, 1994.

[6] J. Contreras, R. Espinola, F.J. Nogales, and A. J. Conejo, "ARIMA models to predict next-day electricity prices", IEEE Transactions on Power System, 18(3):1014 - 1020, 2003.

[7] E. Diaconescu, "The use of narx neural networks to predict chaotic time series", WSEAS Trans. Comp. Res., 3(3):182 - 191, March 2008.

[8] L. A. Daz-Robles, J. C. Ortega, J. S. Fu, G. D. Reed, J. C. Chow, J. G. Watson, and J A.M. Herrera, "A hybrid ARIMA and artificial neural networks model to forecast particulate matter in urban areas: The case of Temuco, Chile", Atmospheric Environment, 42(35):8331 - 8340, 2008.

[9] Google-cluster-usage-trace: http://code.google.com/googleclusterdata.

[10] S. Haykin. Neural Networks: A Comprehensive Foundation. Prentice Hall PTR, 1998.

[11] K.S. Jeong, K.D. Kim, K. M. Jong, M.C. Kim, G.J.Joe, "Non-linear autoregressive modeling by temporal recurrent neural networks for the prediction of fresh water phytoplankton dynamics", Ecological Modeling, 211(3-4):292 - 300, January 2008.

[12] M. Khashei and M. Bijari, "An artificial neural network (p,d,q) model for time series forecasting", Expert Syst. Appl. 37(1):479 - 489., January, 2010.

[13] Y.H.Liang, "Combining seasonal time series arima method and neural networks with genetic algorithm for predicting the production value of the mechanical industry in Taiwan," Neural Computing and applications, 18(7):833 - 841, 2009.

[14] J.M.P. Menezes, Jr. and G. A. Barreto. Long-term time series prediction with the narx network: An empirical evaluation. Neurocomputing., 71(16-18):3335 - 3343, October, 2008.

[15] P.C Nayak, K.P Sudheer, D.M Rangan, and K.S Ramasaatri, "A neuro-fuzzy computing technique for modeling hydrological time series", Journal of Hydrology, 291(12):52 - 66, 2004.

[16] O. Opeyemi and E. O. Justice, "Development of neuro-fuzzy system for early prediction of heart attack", International Journal of Information Technology and Computing (IJITCS), 4(9):22- 28, 2012.

[17] L.H. Tsoukalas and R.E. Uhrig. Fuzzy and neural approaches in engineering, Wiley, 1997.

[18] W.C. Wang, K.C. Chau, C.T. Chengm and L. Qiu, "A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series", Journal of Hydrology, 374(34):294 - 306, 2009.

[19] J.P. Yeh and Y.C Chang, "Comparison neural network and adaptive neuro-fuzzy inference system for forecasting chaotic traffic volumes", Journal of Intelligence Learning System and Applications, 4(4):247 - 258, 2012.

Developing an Automated Bangla Parts Of Speech Tagged Dictionary

Sabir Ismail, M. Shahidur Rahman, Md. Abdullah Al Mumin
Department of Computer Science and Engineering
Shahjalal University of Science and Technology
sabir.ismail01@gmail.com, rahmanms@sust.edu, mumin-cse@sust.edu

Abstract— **This paper develops an algorithm for making an automated Bangla Parts Of Speech (POS) tagged dictionary. Natural Language Processing is one of the most vigorous research areas of computer science. It enables to communicate and retrieve information form computer based system more effectively and efficiently. Researches on Bangla language processing have started long back. However, this research area still suffers from resource scarcity. A POS tagged corpus is a cardinal element for language processing. POS tagging is the process of categorizing a particular word to a particular part of speech or syntactic category. In Bangla, we do not have any large POS tagged dictionary. In this paper we develop an automated way to make a POS tagged dictionary of Noun, Verb and Adjective. Initially, a suffix (or postfix) list is created manually for Bangla language. Based on this suffix list the POS tagged dictionary is developed. The proposed algorithm is evaluated using a paragraph consisting of manually tagged 10,000 words with 11 tags. We found that POS tagging is obtained more accurately for Verb than Noun and Adjective.**

Keywords— Bangla Language Processing, Parts Of Speech Tagging, Machine Learning, Bangla Corpus.

I. INTRODUCTION

Natural Language Processing [1-2] comprises large area of researches like information retrieval, document categorization, machine translation, question answering, human-computer interaction, automatic summarization, coherence resolution, discourse analysis, morphological segmentation, naming entity recognition and so on. Natural language processing starts with separating each word from a sentence. These individual words are further processed for more advance analysis. The mere individual words do not contain enough information for further processing. POS tag of word helps profound parsing of text, semantic processing and morphological analysis. There are various ways for POS tagging such as supervised, semi supervised or unsupervised methods. In supervised methods there are fixed rules for tagging. Unsupervised methods used machine learning techniques like Hidden Markov Model (HMM), Maximum Likelihood Estimation and Maximum Entropy (ME) based tagging. Semi-supervised methods use both supervised and unsupervised methods. In this paper, we propose a semi-supervised method to develop POS tag dictionary for noun, verb and adjectives. There is no large size POS tag dictionary currently available for Bangla. This POS tag dictionary can be employed for word categorization, word sense disambiguation, morphological analysis of verb and noun, analyse suitable POS tag set for Bangla. It can be used as training corpus for unsupervised models as well.

The rest of the paper is organized as follows. We start with an overview of previous work on Bangla corpus and Bangla POS tagging in Section II. Then we discuss Bangla language and use of parts-of-speech in Bangla grammar in Section III. Section IV describes the proposed algorithm and Section V analyses the results. Finally we conclude the paper in Section VI.

II. RELATED WORKS

Bangla language is the 7th most spoken language [3] in the world. Compared to other languages like English, Chinese, German, and French, Bangla has scarcity of large corpus. Different types of corpuses are available for those languages. Unfortunately, effort for building a rich POS tagged dictionary of Bangla has just started. The Centre for Research on Bangla Language Processing of BRAC University has conducted some works. They proposed a way to implement a corpus by collecting data form online resources [4]. They implemented POS tagger based on HMM, n-gram and Brill's tagger [5]. The result is analized with a small corpus of 5000 words giving accuracy of only 55%. Since Bangla language is a head-final specifier-initial language, another study shows that backward n-gram gives better result than forward n-gram [6]. Ansari, Selim and Iqbal [7] also make some statistical analysis on Bangla Academy Dictionaries corpus. Their corpus also contains only 150 tagged words. Hammad Ali [8] developed Baum-Welch trained HMM tagger for the Bangla POS tagging. His aim was to check whether rule based taggers was working better than stochastic taggers. The author used 41 tag sets and 50,000 tokens and outcome shows that rule based tagger is more effective than tagger. Dandapat, Sandipan, Sudeshna and Anupam [9] build an automatic POS tagger using HMM, ME based stochastic taggers. They also used morphological and contextual information of the words to improve performance. They summarized that morphological features are helpful to develop Bangla POS tagger when tagged resources are limited. Chakrabarti, Ekbal, Asif, Rejwanul and Sivaji [10] proposed a Bangla POS tagger using the statistical Conditional Random Fields and they also conclude that different contextual information of the words

along with the variety of features are helpful in predicting POS tag. They also developed a ME based tagger [11] with a training corpus of 72,341 words and 26 tag sets. They compare result with HMM and ME outperforms the HMM based tagger. They also added that lexicon, naming entity recognizer and different word suffixes are effective in handling unknown word tagging problems. Another attempt by Ekbal, Asif and Sivaji [12] is lexicon development and POS tagging by HMM with approximately 31,190 word forms and 5967 manually tagged corpuses. They achieved 83.04% accuracy and accuracy increased up to 91.6% with inclusion of different techniques adopted for handling unknown words. Dandapat and Sandipan [13] used ME based static parser for Bangla, Hindi and Telugu, and also used morphological analyzer for Bangla that led better result for Bangla tagger than Hindi and Telugu. Another layer based POS tagging was attempted by Chakrabarti, Debasri and Pune [14], they used a rule based tagger and 4 layers of chunking to handle multi-verb exception and reduced ambiguous tagging. Here contextual and morphological cue plays vital role in different layers. All of them used a manually created tagged corpus for training semi-supervised and unsupervised model or used rule based tagger, but no attempt is made to develop a large tagged dictionary.

III. BANGLA LANGUAGE

Bangla Language arose from eastern Middle Indo-Arayn dialects of the Indian subcontinent, which is now the main language of Bangladesh and some western parts of India. Bangla language has rich grammatical rules including Parts-of-speech (পদ), Case (কারক), Suffix (বিভক্তি), Prefix (উপসর্গ), Number (বচন), Gender (লিঙ্গ), Person (পুরুষ), Tense (কাল), Primary Suffix (কৃত প্রত্যয়) and Secondary Suffix (তদ্ধিত প্রত্যয়) [17-22]. The structure of Bangla language is quite different form English language. Bangla words are highly inflectional particularly for verb and noun. A Bangla verb has more than 200 inflectional forms based on tense, aspect, mood and person. Mostly those words are formed by adding suffix with some exceptions. Same is true for noun. Based on case, gender and number, noun also has several inflectional forms. This makes vast variation of Bangla words. However, this highly inflection of word has also some benefits. We use this inflection of words to build POS tag dictionary.

POS tagging is the process of tagging every word in a text to a particular parts of speech based on relation with neighboring words and phrase. English language has mainly 8 parts of speech: Noun, Pronoun, Verb, Adverb, Adjective, Preposition, Conjunction and Interjection. However, number of tags varies from corpus to corpus when tagging POS. English language, for example, has several corpuses where every corpus has its own tag set. The Brown Corpus has 87 tag set, the Penn Treebank has 45 and the C7 has 146 tag set [15]. Bangla language has only five parts of speech: Noun, Pronoun, Adjective, Verb and Conjunction. Due to absence of large corpus, the number of tag set for Bangla is not well defined. In this paper, we work only with Noun, Verb and Adjective. The *root word* is a primary lexical unit of a word, which contains most significant information about semantic meaning and cannot be reduced to smaller part. Sometimes, the term *root word* is also used to describe the word with its inflectional endings discarded, but with its lexical endings in place. This *root word* is also called lemma or stems and inflectional endings are called suffix. In Bangla language a word can have more than one suffix or inflectional endings added to it. Our proposed method uses this suffix or inflectional ending to determine the POS tag.

IV. DEVELOPING BANGLA POS TAG DICTIONARY

As discussed earlier, Bangla words are highly inflectional. Almost every word has more than one inflectional form. We use this inflectional form to determine the POS tag. First, a list of Bangla suffix [17-22] is created from various resources on Bangla grammars. This list contains both the Primary Suffix (কৃত প্রত্যয়) which are added with verb and Secondary Suffix (তদ্ধিত প্রত্যয়) which are added with noun. This suffix list also contains some suffixes which are used to change number (বচন), gender (লিঙ্গ) and some suffixes are used to change POS tag of a word. We listed about 500 suffixes. Then we collect large number of words from various sources, like online newspapers, blogs and other web sites [23-27]. Primarily, total number of words collected are about 1000000. After eliminating the duplicate words the number is reduced to 320443 unique words. Then the words are sorted according to the Unicode values of Bangla characters [16]. As data structure, a hash table (also called hash map) is used that map the keys to values for storing data.

A. The Proposed Algorithm:

This algorithm is described here for a number of words and suffixes as given in Table 1 and Table 2.

1. Initially we introduced three hash tables: Hash map-1, Hash map-2 and Hash map-3. Hash map-1 (Key: root word, Value: suffix) contains the root word as *Key* and suffixes as the *Values*. Hash map-2 (Key: root word, Value: inflectional form) stores the root word as *Key* and the inflectional form of that root word as *Values*. Hash map-3 (Key: suffix, Value: root word) contains suffix as the *Key* and the root words as *Values*. Initially all the hash maps are set empty.

TABLE 1. INITIAL WORD LIST

Word List
অংশ,অংশই,অংশও,অংশকে,অংশগুলির,অংশগুলো,অংশগুলোও,অংশগুলোতে, অংশগ্রহণ,অংশগ্রহণও,অংশগ্রহণকারী,অংশগ্রহণকারীদের,অংশগ্রহণকারীর, অংশগ্রহণকারীরা,অংশগ্রহণকারীরাই,অংশগ্রহণকে, অংশগ্রহণরত,অংশগ্রহণে,অংশগ্রহণের

TABLE 2. SUFFIX LIST

Suffix list:
া, ি, ই, ও, কে, গুলি, গুলির, গুলো, গুলোও, গুলোতে, দের, র, রা, রাই, কে, ত, ে, ের

2. Then for every word in the word list we check if the starting of a word matches with any root word stored in Hash map-1 (Key: root word, Value: suffix) then we chop the root part from that word and retrieve rest of the characters.
3. If rest of the characters are found in the suffix list that indicates the word is generated from that root word, then this

word is added in Hash map-1, Hash map-2 and Hash map-3. The example 1 as given below explains the case.

4. If a word matches with the root word more than once then the longest match is considered. Again, if a word matches with the root word but there are some other words that starts with this word, then this word itself is a root word. Explanation is given in example 2.

Example 1: অংশ is the first word in our word list, so this is a root word which is stored in hash map. Table 3 shows the content of hash map.

TABLE 3. CONTENT OF THE HASH MAPS AFTER PROCESSING THE FIRST WORD অংশ

Hash map-1 (Key: root word, Value: suffix)	**Hash map-2 (Key: root word, Value: inflectional form)**	**Hash map-3 (Key: suffix, Value: root word)**
Key: অংশ, Value: root	Key: অংশ, Value: root	Key: root, Value: অংশ

The next word is অংশই, now first we check in the Hash map-1 (Key: root word, Value: suffix) if there is any word which is a substring and also a prefix of অংশই. In this case, অংশ matches both the conditions. Then we chop rest of the characters from অংশই which is ই and check whether it is in suffix list. In this case it is found and অংশই is asserted as an inflectional form of অংশ and the hash maps are updated as shown in Table 4.

TABLE 4. CONTENT OF THE HASH MAPS AFTER PROCESSING FIRST TWO WORDS অংশ, অংশই

Hash map-1 (Key: root word, Value: suffix)	**Hash map-2 (Key: root word, Value: inflectional form)**	**hash map3 (Key: suffix, Value: root word)**
Key: অংশ, Value: root, ই	Key: অংশ, Value: root, অংশই	Key: root, Value: অংশ Key: ই, Value: অংশ

Example 2: Let us now assume the word অংশগুলো in the word list. Before we start processing the word অংশগুলো, the words অংশ, অংশই, অংশও, অংশকে, অংশগুলির are already processed and Table 5 shows the content of the hash maps.

TABLE 5. CONTENT OF HASH MAPS BEFORE PROCESSING THE WORD অংশগুলো

Hash map-1 (Key: root word, Value: suffix)	**Hash map-2 (Key: root word, Value: inflectional form)**	**Hash map-3 (Key: suffix, Value: root word)**
Key: অংশ, Value: root, ই, ও, কে, গুলির	Key: অংশ, Value: root, অংশই, অংশও, অংশকে, অংশগুলির	Key: root, Value: অংশ Key: ই, Value: অংশ Key: ও, Value: অংশ Key: কে, Value: অংশ Key: গুলির, Value: অংশ

5. Now অংশগুলো matches with root word অংশ and গুলো is a valid suffix, but there are some more words (অংশগুলোও, অংশগুলোতে) also start with অংশগুলো. In this case অংশগুলো is considered as a root word. Table 6 shows content of the hash map after processing the word অংশগুলো.

TABLE 6. CONTENT OF THE HASH MAPS AFTER PROCESSING THE FIRST WORD অংশগুলো

Hash map-1 (Key: root word, Value: suffix)	**Hash map-2 (Key: root word, Value: inflectional form)**	**Hash map-3 (Key: suffix, Value: root word)**
Key: অংশ, Value: root, ই, ও, কে, গুলির Key: অংশগুলো, Value: root	Key: অংশ, Value: root, অংশই, অংশও, অংশকে, অংশগুলির Key: অংশগুলো, Value: root	Key: root, Value: অংশ, অংশগুলো Key: ই, Value: অংশ Key: ও, Value: অংশ Key: কে, Value: অংশ Key: গুলির, Value: অংশ

Table 7 shows the content of the hash map after processing all the words in the wordlist.

TABLE 7. CONTENT OF THE HASH MAPS AFTER PROCESSING ALL WORDS IN THE WORDLIST

Hash map-1 (Key: root word, Value: suffix)	**Hash map-2(Key: root word, Value: inflectional form)**	**Hash map-3 (Key: suffix, Value: root word)**
Key: অংশ, Value: root, ই, ও, কে, গুলির Key: অংশগুলো, Value: root, ও, তে Key: অংশগ্রহণ, Value: root, ও, কে, ে, ের Key: অংশগ্রহণকারী, Value: root, দের Key: অংশগ্রহণকারীর, Value: root, া, াই Key: অংশগ্রহণরত, Value: root	Key: অংশ, Value: root, অংশই, অংশও, অংশকে, অংশগুলির Key: অংশগুলো, Value: root, অংশগুলোও, অংশগুলোতে Key: অংশগ্রহণ, Value: root, অংশগ্রহণও,অংশগ্রহণকে, অংশগ্রহণে, অংশগ্রহণের Key: অংশগ্রহণকারী, Value: root, অংশগ্রহণকারীদের Key: অংশগ্রহণকারীর, Value: root,অংশগ্রহণকারীরা, অংশগ্রহণকারীরাই Key: অংশগ্রহণরত, Value: root	Key: ের, Value: অংশগ্রহণ Key: ে, Value: অংশগ্রহণ Key: ই, Value: অংশ Key: ও, Value: অংশ, অংশগুলো, অংশগ্রহণ Key: কে, Value: অংশ, অংশগ্রহণ Key: গুলির, Value: অংশ Key: তে, Value: অংশগুলো Key: দের , Value: অংশগ্রহণকারী Key: া, Value: অংশগ্রহণকারীর Key: াই, Value: অংশগ্রহণকারীর Key: root, Value: অংশ, অংশগুলো, অংশগ্রহণ, অংশগ্রহণকারী, অংশগ্রহণকারীর, অংশগ্রহণরত

6. We analyse data stored in the Hash map-3 (Key: suffix, Value: root word) and it is revealed that some suffixes are very reliable candidate for POS tagging a word. For example, গুলি, টি, তে suffixes are secondary suffix and are added with only noun. Similarly ছিল, িব, িতেছিল suffix are primary suffix and are added with only verb. We, therefore, categorize the suffixes as verb, noun and adjective. We tagged nouns as 'N', verb as 'V', adjective as 'A' and some suffixes are not suitable for detecting POS tag which are tagged as 'U' as shown in Table 8.

TABLE 8: POS TAGGED SUFFIX LIST

Suffixes with POS tag					
িয়াছ V	কি U	গুলিতে N	তাম V	তিছি V	তুত V
িয়াছি V	কী U	গুলিন N	তার V	তিছেন V	তুম V
িয়াছিলাম V	কুল U	গুলির N	তি U	তিস V	তাছেন V
কারী N	তা A	গুলো N	তিও U	কুলের N	গুলি N
তর A	তাছি V	তম A	তাও U	কে N	গুলিকে N

Then we create two more hash maps.

- Hash map-4 (Key: root word, Value: POS tag) from Hash map-1 (Key: root word, Value: suffix) , by replacing suffixes with POS tag of Table 8 as shown in Table 9

- Hash map-5 (Key: root word, Value: inflectional form + POS tag) from Hash map-2 (Key: root word, Value: inflectional form), by replacing suffixes with POS tag of Table 8 as shown in Table 10.

TABLE 9. HASH MAP WITH POS TAG

Hash map-1 (Key: root word, Value: suffix)	Hash map-4 (Key: root word, Value: POS tag)
Key: অংশ, Value: root, ই, ও, কে, গুলির Key: অংশগুলো, Value: root, ও, তে Key: অংশগ্রহণ, Value: root, ও, কে, ে, েের Key: অংশগ্রহণকারী, Value: root, দের Key: অংশগ্রহণকারীর, Value: root, া, াই Key: অংশগ্রহণরত, Value: root	Key: অংশ, Value: root, U, U, N, N Key: অংশগুলো, Value: root, U, N Key: অংশগ্রহণ, Value: root, U, N, U, N Key: অংশগ্রহণকারী, Value: root, N Key: অংশগ্রহণকারীর, Value: root, U, N Key: অংশগ্রহণরত, Value: root

TABLE 10. HASH MAP WITH POS TAG AND INFLECTIONAL FORM

Hash map-2 (Key: root word, Value: inflectional form)	Hash map-5 (Key: root word, Value: inflectional word and POS tag)
Key: অংশ, Value: root, অংশই, অংশও, অংশকে, অংশগুলির Key: অংশগুলো, Value: root, অংশগুলোও, অংশগুলোতে Key: অংশগ্রহণ, Value: root, অংশগ্রহণও,অংশগ্রহণকে, অংশগ্রহণে, অংশগ্রহণের Key: অংশগ্রহণকারী, Value: root, অংশগ্রহণকারীদের Key: অংশগ্রহণকারীর, Value: root, অংশগ্রহণকারীরা, অংশগ্রহণকারীরাই Key: অংশগ্রহণরত, Value: root	Key: অংশ, Value: root, অংশই U, অংশও U, অংশকে N, অংশগুলির N Key: অংশগুলো, Value: root, অংশগুলোও U, অংশগুলোতে N Key: অংশগ্রহণ, Value: root, অংশগ্রহণও U, অংশগ্রহণকে N, অংশগ্রহণে U, অংশগ্রহণের N Key: অংশগ্রহণকারী, Value: root, অংশগ্রহণকারীদের N Key: অংশগ্রহণকারীর, Value: root, অংশগ্রহণকারীরা U, অংশগ্রহণকারীরাই N Key: অংশগ্রহণরত, Value: root

7. We eliminate multiple occurrences of same tag (e.g. multiple 'N' or 'U') form the Hash map-4 (Key: root word, Value: POS tag) as shown in Table 11. We remove 'U' tag, because our analysis on suffix list reveals that these suffixes cannot determine POS tag and they do not also change POS tag of a root word. Thus these inflectional words have same POS tag as root word.

TABLE: 11 HASH MAP WITH POS TAG FOR ROOT WORD

Hash map-6 (Key: root word, Value: tag)	
অংশ [root, N] অংশগুলো [root, N] অংশগ্রহণ [root, N]	অংশগ্রহণকারী [root, N] অংশগ্রহণকারীর [root, N] অংশগ্রহণরত [root]

Finally we regenerate all the words from Hash map-2 (Key: root word, Value: inflectional form) and Hash map-6 (Key: root word, Value: tag). Table 12 shows the POS tag dictionary.

TABLE 12. FINAL POS TAGGED DICTIONARY

POS tagged Dictionary		
অংশ [N] অংশই [N] অংশও [N] অংশকে [N] অংশগুলির [N] অংশগুলো [N] অংশগুলোও [N]	অংশগুলোতে [N] অংশগ্রহণ [N] অংশগ্রহণও [N] অংশগ্রহণকে [N] অংশগ্রহণে [N] অংশগ্রহণের [N]	অংশগ্রহণকারী [N] অংশগ্রহণকারীদের [N] অংশগ্রহণকারীর [N] অংশগ্রহণকারীরা [N] অংশগ্রহণকারীরাই [N] অংশগ্রহণরত [U]

V. EXPERIMENT AND RESULT ANALYSIS

We have collected more than 10,00,000 words from online Bangla newspapers, blogs and other Bangla web sites and 3,20,443 words are extracted as unique. The proposed technique successfully tagged 1,34,749 nouns, 11,067 verbs, and 8,435 adjectives. Some POS has very small amount of words, compare with noun or verb. Example: pronoun, conjunction, interjection, *Wh* question, negative, preposition has small amount of words. So our algorithm do not try to predict those tag set, rather we make a list of words having those tag set. The number of words are increasing every day that will further improve the accuracy of POS tag. We evaluate the proposed algorithm using a paragraph consisting of manually tagged 10,000 words with 11 tags as shown in Table 13.

TABLE 13. POS TAG LIST FOR TAGGED CORPUS

POS	Tag	POS	Tag
Noun	NN	Adverb	RB
Pronoun	PRP	Adjective	JJ
Verb	VB	Postposition	PSP
Auxiliary Verb	VA	Conjunction	coNN
Wh question	WH	Interjection	INT
Negative	NEG	Unknown	U

Table 14 shows the Confusion Matrix resulted from the experiment where the horizontal tags represent the actual and the vertical tags represent the obtained tags. We work for Noun (NN), Verb (VB), Adjective (JJ), unknown tag (U) and others are tagged as (O). A number in a particular cell represents the number of tags obtained using the proposed method with respect to the actual tag. For example 3201 NN tags are determined successfully 2589 times.

TABLE 14. CONFUSION MATRIX OBTAINED USING A MANUALLY TAGGED CORPUS OF 10000 WORDS

		Actual POS tag											
POS tag obtained using proposed algorithm		**VB (2296)**	**WH (51)**	**coNN (125)**	**NEG (135)**	**NN (3201)**	**U (0)**	**VA (763)**	**INT (86)**	**RB (312)**	**PRP (1298)**	**JJ (966)**	**PSP (0)**
	VB	2036	0	0	0	224	0	711	0	6	0	5	0
	NN	219	0	0	0	2589	0	14	3	5	12	513	0
	JJ	9	0	1	0	39	0	8	0	0	0	311	0
	U	3	2	2	2	26	767	4	77	45	29	81	0
	O	29	49	122	133	123	0	26	6	256	1257	56	0

Table 15 shows the accuracy of the proposed algorithm.

TABLE 15. OVERALL PERFORMANCE

Tag	True Positive	False Positive	False Negative
Verb (VB)	2747	235	260
Noun (NN)	2589	766	412
Adjective (JJ)	311	57	655

Table 15 shows that POS tagging is obtained more accurately for both Verb (VB) and Auxiliary Verb (VA) (2036+711= 2747). But for Noun (NN), False Positive and for Adjective (JJ), False Negative is high. This is because Noun and Adjective tag are interchangeable for a word based on its contextual meaning and surrounding words. Here we do not consider surrounding words which are usually used in machine learning based tagging methods.

VI. CONCLUSION

In this paper, we have developed method for tagging noun, verb and adjective. The proposed algorithm is evaluated using a paragraph consisting of manually tagged 10,000 words with 11 different tags. We found that POS tagging is obtained more accurately for Verb than Noun and Adjective.

Resource is prerequisite to any research. Unfortunately, Bangla language still suffers from scarcity of resource. The resulted POS tag dictionary can play a vital role for future research on Bangla language processing specially on POS tagging and making a canonical tag set. Unsupervised model like HMM needs large tagged corpus for training data. The POS tag dictionary can serve this purpose.

REFERENCES

[1] A free encyclopedia built collaboratively using wiki software: https://en.wikipedia.org/wiki/Natural_language_processing :Last accessed 14th November, 2013.

[2] Natural Language Processing: http://research.google.com/pubs/NaturalLanguageProcessing.html : Last accessed 14th November, 2013.

[3] "Statistical Summaries". Ethnologue. Retrieved 2013-07-31. : Last accessed 14th November, 2013.

[4] Sarkar, Asif Iqbal, Dewan Shahriar Hossain Pavel, and Mumit Khan."Automatic Bangla corpus creation". Center for research on Bangla language processing (CRBLP), BRAC University, 2007.

[5] Hasan, Fahim Muhammad, Naushad UzZaman, and Mumit Khan. "Comparison of different POS Tagging Techniques (N-Gram, HMM and Brill's tagger) for Bangla." Advances and Innovations in Systems, Computing Sciences and Software Engineering. 2007.

[6] Khan, Naira, et al. "History (Forward N-Gram) or future (Backward N-Gram)? Which model to consider for N-Gram analysis in Bangla?" 2006.

[7] MD. Abdul Awal Ansary,Mohammad Reza Selim, Muhammad Zafar Iqbal. "Bangla Academic Dictionaries (BAD) Corpus with some Applications and Statistical Analysis". Journal of Emerging Trends in Computing and Information Sciences. VOL. 3, NO.11 Nov, 2012.

[8] Ali, Hammad. "An unsupervised parts-of-speech tagger for the bangla language." Department of Computer Science, University of British Columbia, 2010.

[9] Dandapat, Sandipan, Sudeshna Sarkar, and Anupam Basu. "Automatic part-of-speech tagging for Bengali: An approach for morphologically rich languages in a poor resource scenario." Proceedings of the 45th Annual Meeting of the ACL on Interactive Poster and Demonstration Sessions. Association for Computational Linguistics, 2007.

[10] Chakrabarti, Ekbal, Asif, Rejwanul Haque, and Sivaji Bandyopadhyay. "Bengali part of speech tagging using conditional random field." Proceedings of Seventh International Symposium on Natural Language Processing, 2007.

[11] Ekbal, Asif, Rejwanul Haque, and Sivaji Bandyopadhyay. "Maximum Entropy Based Bengali Part of Speech Tagging." RCS Journal 33, pp.67-78, 2008.

[12] Ekbal, Asif, and Sivaji Bandyopadhyay. "Lexicon Development and POS Tagging Using a Tagged Bengali News Corpus." FLAIRS Conference, 2007.

[13] Dandapat, Sandipan. "Part of specch tagging and chunking with maximum entropy model." Proceedings of the IJCAI Workshop on Shallow Parsing for South Asian Languages, Hyderabad, India, 2007.

[14] Chakrabarti, Debasri, and Pune CDAC. "Layered Parts of Speech Tagging for Bangla." Language in India, www. languageinindia. com, Special Volume: Problems of Parsing in Indian Languages, 2011.

[15] Daniel Jurafsky, James H. Martin – "Speech and Language Processing"(2nd Edition), Page: 8; Chapter: 5, 2008.

[16] Amin, Md Ruhul, Asif Mohammed Samir, Madhusodan Chakraborty, and Mahfuzur Rahaman. "An Efficient Unicode based Sorting Algorithm for Bengali Words." *International Journal of Computer Applications* 24, no. 7, 2011.

[17] ড.আলাউদ্দিন আল আজাদ, ড.মনন অধিকারী, রুহুল আমীন বাবুল - "উচ্চতর বাংলা ভাষারীতি" লেকচার পাবলিকেশন, ২০০৮ ইং.

[18] নবম-দশম শ্রেনী ২০০০- জাতীয় শিক্ষাক্রম ও পাঠ্যপুস্তক বোর্ড-"বাংলা ব্যাকরন",২০১১ ইং.

[19] ড.মুহম্মদ শহীদুল্লাহ - "বাংলা ব্যাকরন", মাওলা বার্দ্রাস, ১৪১০ বাং.

[20] ড.হুমায়ুন আজাদ - "বাক্য তত্ত্ব", আগামী প্রকাশনী, ২০১০ ইং.

[21] মহাম্মদ দানীউল হক - "ভাষা বিজ্ঞানের কথা" , মাওলা বার্দ্রাস, ২০১১ ইং.

[22] উদয়কুমার চক্রবর্তী - "বাংলা পদগুছের সংগঠন" , দেজ পাবলিশিং কলকাতা ২০০৪ ইং.

[23] মুহম্মদ জাফর ইকবাল - "আমার বন্ধু রাশেদ", কাকলী প্রকাশনী, ১৯৯৮ ইং.

[24] বাংলা ব্লগ: http://www.somewhereinblog.net/ : Last accessed 14th November, 2013.

[25] Bangla Newspaper, প্রথম আলো: http://www.prothom-alo.com/ : Last accessed 14th November, 2013.

[26] Bangla Newspaper, কালের কন্ঠ: http://www.kalerkantho.com/ : Last accessed 14th November, 2013.

[27] অনলাইন বাংলা নিউজ : http://www.rtnn.net/ : Last accessed 14th November, 2013.

Structured Web Search in Small Domain

Mohiuddin Abdul Qader, Sajjadur Rahman and Masud Hasan
Department of Computer Science and Engineering
Bangladesh University of Engineering and Technology, Dhaka-1000, Bangladesh
Email: mohiuddin.a.qader@gmail.com, {sunnysajjad, masudhasan}@cse.buet.ac.bd

***Abstract*—In this paper our aim is to bridging the gap between machine interpretation of search queries in search engines and actual user queries. Our primary concentration is to find a way to represent diverse search result pages in a structured way to reduce search query complexity and satisfy the need of diverse people. We first propose a traditional small scale search engine, called *BuetSearch*, for Bangladesh University of Engineering and Technology (BUET) domain, which has been designed by combining the *PageRank* algorithm and a proposed ranking and weighting scheme for Information Retrieval scoring. Then we propose a *K*-nearest neighbor algorithm based document classification algorithm for structured representation of search results. The results show that web search in small domain has merit over traditional large scale search engines and can be optimized to provide excellent results. With the application of structured approach, web search can be significantly simplified and the inconvenience of the users can be reduced.**

I. Introduction

Typical search engines exhibit good performance in document-finding, but their search process is limited to extracting documents based on relevance-ranking queries. They cant predict the user's expectation from a short query. There is often a discrepancy between the search results and the actual user queries. This mismatch arises due to the difference between the way document creators describes information and the way users try to retrieve them. Therefore, recognizing the alternative ways people use to reference an entity, is crucial for web search. For example: let's search *U*2 in typical search engines, i.e. *Google*/*Bing* etc. and see what comes up. We all know *U*2 is one of the greatest rock bands of all time, so does *Google*/*Bing*. That's why *Google*/*Bing* will provide you millions of search results on *U*2-the band. But what if I was not searching *U*2-the music band, rather *U*2-the short drama or *U*2-the aircraft or *U*2-the mathematical function. Is there any way for search engines to predict the user's expectation from such short query? Can we map user's search query to some sort of categorization of data prevailing in World Wide Web? With vast amount of data in WWW, is such mapping feasible? The growing demand of structured data has led us to the idea of breaking it down, let's say; apply divide and conquer approach by subdividing the web in scalable domains for simplicity and deploy structured and categorized representation of search results to deal with query complexity. We built our search engine *BuetSearch* as a prototype for this idea where the relevant documents obtained against user query are classified in multiple predefined categories.

Users nowadays, are posing more complex queries to search engines. Interpreting such complex queries require certain degree of intelligence. So, search engine developers are concentrating on keeping pace with this evolutionary nature of user query at every instance and coming up with delicate but efficient strategies. In recent years, Google has come up with a number of projects with special importance to structured web search, e.g. Google Squared, Google Wonder Wheel, Google Timeline etc. These projects are currently not live. Faceted search engines [6] have been developed both for academic and commercial purpose, e.g. *Eccellio*, *Flamenco* etc.

A simpler solution to the complex Web of data is to localize the information. Profiling of user tendencies on a local basis may provide a basis for personalized search. The ranking criteria can be modified likewise. Such efforts have already been made, e.g, *Ansearch* and *Baidu*. *Ansearch* uses popularity as the key factor in its algorithm to calculate search result rankings. It features two unique services not found on other search engines - the top 100 most popular websites and the top 100 most popular keywords. *Baidu*, on the other hand, focuses on powering the best technology, optimized for up-to-date local tastes and preferences. Their deep understanding of Chinese language and culture is central to their success and this kind of knowledge allows them to tailor the search technology for their users' needs.

In our approach, at first we have developed *BuetSearch* as a traditional search engine with *PageRank* [7] as ranking scheme, Vector Space Model [8] as Information Retrieval (IR) Model and $tf - idf$ document weighting [1] as the IR model. Then, we shift our concentration to structured web search in small domain. Since the paper concentrates on structured web search in small domain, less efforts were given in developing a highly efficient traditional search engine. Our developed structured prototype can be easily used on top of any traditional search engine and thus, provide a structured web search interface. Searching in small domain has an added advantage: the scope is a fixed domain. So the search patterns of users, the web usage culture of the domain and several other factors influence the query evaluation process of the search engine. To train the search engine for multi-category classification, a supervised learning method called *K*-nearest neighbour [10] [9] based document classification algorithm is proposed and implemented.

Keeping the importance of human computer interaction in focus, we present our search results in a different

978-1-4799-3497-3/13/$31.00 ©2013 IEEE

way. For a given query, the search results are obtained, then classified in two most relevant categories and presented in a split screen format where each segment of the screen shows search result of a different category.

II. *BuetSearch* System Architecture

The high level architecture of *BuetSearch* is described in this section. The *BuetSearch* database consists of 10 tables (schema omitted from here due to space restraints). The most notable of them are: *URLList*, *PrimaryStorage*, *AnchorStorage*, *HitList*, *WordStorage*, *SortedPageList*, *SourceTable* and *CategorizedStorage*. Since the system is implemented in a small domain (BUET), it contains a single *Crawler* to crawl (download) Web Pages. The *URLList* stores the list of URLs in the domain. The *Crawler* fetches an URL from the *URLList* and crawls the corresponding Web page from BUET server. The crawled page is stored in a directory in the *BuetSearch* server, *PageStore*. For the time being, we have crawled only HTML pages. The *Crawler* has crawled about 1900 Web pages. We have performed Breadth First crawling. The Indexer fetches

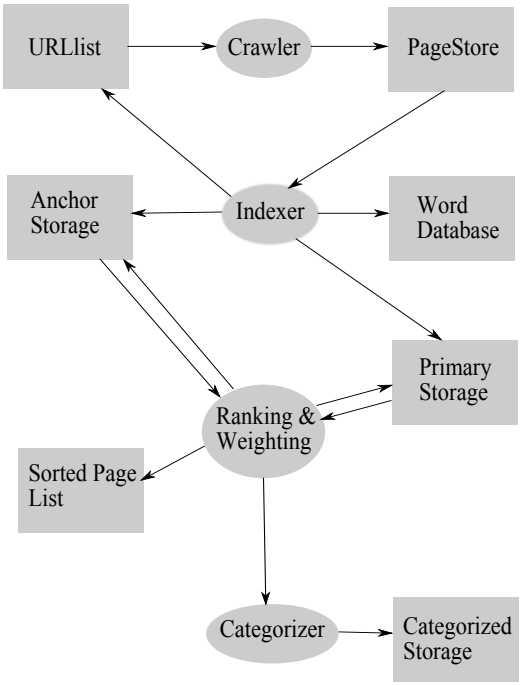


Fig. 1. Architecture of *BuetSearch*

each HTML page from the *PageStore* and parses the page into various portions. From each Web page it extracts relevant information: Page title, Pain Text, Hyperlinks, Anchor text, number of outbound links etc. Most of the extracted information are stored in *PrimaryStorage*, e.g. Title, Plain Text etc. The anchor text is stored in *AnchorStorage*. From the hyperlinks found in the pages, a source to destination mapping is stored in *SourceTable*. Using the Ranking and Weighting schemes (discussed in Section III), the system then calculates the combined value of *PageRank* and IR score and stores the value for each page in *SortedPageList*. A *Categorizer* is used to categorize the pages in predefined classes/categories and stored on *CategorizedStorage*. The categorization process is discussed in Section IV.

III. Ranking and Weighting Schemes

We have used Google's *PageRank* algorithm [3] [7] as our ranking scheme.

A. Computing IR Score

For computing the Information Retrieval (IR) Score we have used the notion of Vector Space Model [8]. We have implemented nearly the same idea in our IR scoring algorithm. The main purpose of idf weighting is to remove any unnecessary *stopwords* from documents or search query. Procedure **ListDifference(**$list1, list2$) is used to remove the *stopwords* from a given document.

Following table shows the weighting scheme we used in our system to calculate tf weight:

TABLE I. Table for assigning weight to various features of a Web Page

Feature	Feature Contains	% of tf weight
Page Title	(a) Entire Query String	0.8
	(b) Reverse Query String	0.7
	(c) Only 1st Query Word	0.6
	(d) Only 2nd Query Word	0.5
Plain text	(a) Space (' ') Separated Query	0.5
	(b) Both the Query Words separately	0.3
	(c) Only 1st Query Word	0.2
	(d) Only 2nd Query Word	0.15

The tf weight of a document d is given by-

$$tf_d = \sum_{i=1}^{M} n_{FC_i} \times weight_{FC_i} \tag{1}$$

where n_{FC_i} is the frequency of i-th feature characteristic in d and $weight_{FC_i}$ is the fractional weight of tf value of i-th feature characteristic. Here, the number of features considered, $M = 2$.

B. Query Evaluation Process in BuetSearch

Whenever user submits a query to the system, the query evaluation process is activated and it performs a series of subroutines to present SERPs to the user. First, the system calculates IR score of each document in *PrimaryStorage* against the user query and then combines the result with *PageRank* to obtain a measure of relevance of documents against the user query.

At first, *title* of each page in *PrimaryStorage* is searched to find if any of them contains the entire *query* or any word of the *query*. If such pages are found then their *titleScore* is assigned using table I. After calculating *titleScore*, we calculate the relevance weight of plain text of pages with respect to *userquery*.

The *titleScore* and *plainTextScore* are combined to calculate tf weight of each page. tf weight is obtained by-

$$tf[p] = titleScore[p] + plainTextScore[p] \times 0.2 \tag{2}$$

Major share of weight is given to *title* than *plainText* (20%) as was discussed in III-A. Finally, the overall relevance of a web page is calculated. User's tendency is to either click the 1st link encountered [2] [4] or at least check the links in first page by reading the text snippet shown by the system and then click on the link that he/she thinks to be relevant. So, we introduce an additional weighting factor *ClickCount* to compute the relevance of a Web page. Relevance weight of a Web page p is measured by-

$$relevanceScore[p] = tf[p] \times 0.5 + ClickCount[p] \times 0.1 + \\ + pageRank[p] \times 0.4 \quad (3)$$

input : userquery
output: List of Relevant Web Pages with their relevance score, *relevanceScore*

query ← `ListDifference`(*userquery*, *stopwords*)
foreach *page p* ∈ *PrimaryStorage* **do**
  titleScore[*p*] ← `TitleScore`(*p*, *userquery*)
  plainTextScore[*p*] ← `PlainTextScore`(*p*, *userquery*)
  $tf[p] \leftarrow titleScore[p] + plainTextScore[p] \times 0.2$
  Assign to *ClickCount*[*p*] number of time page *p* is visited by user clicking on *p*'s link
  Calculate *relevanceScore*[*p*] according to Equation 3
end

Algorithm 1: Query Evaluation in Traditional Search Engine

Here, the weights assigned to tf (50%), click count (10%) and *PageRank* (40%) are based on experimentation.

C. A Sample Scenario

Let, for a given query "$w1$ $w2$", title or text matches were found for 3 pages: $p1$,$p2$ and $p3$. Now we are going to compute *relevanceScore* for the pages by calculating *titleScore*, *plainTextScore*, *ClickCount* and *PageRank*. From the calculation we can show the order in which the pages will be presented to the user.

TABLE II. Calculating tf Score of Web Pages (The feature references correspond to table I)

Page	Title Feature	Plain Text Contains			tf score
		Entire *query*	"w1"	"w2"	
$p1$	(a)	twice	twice	5 times	2.4
$p2$	(c)	–	–	11 times	2.25
$p3$	(d)	once	8 times	6 times	2.8

The corresponding calculation-
$tf[p1] = 0.8 + 2 \times 0.5 + 2 \times 0.3 = 2.4$
$tf[p2] = 0.6 + 11 \times 0.15 = 2.25$
$tf[p3] = 0.5 + 1 \times 0.5 + 6 \times 0.3 = 2.8$

TABLE III. Calculating Relevance Score of Web Pages

Page	tf score	Click Count	*PageRank*	*RelevanceScore*
$p1$	2.4	14	0.39	2.756
$p2$	2.25	11	0.54	2.433
$p3$	2.8	23	0.35	3.84

According to Equation 3 the *relevanceScore* for $p1$, $p2$ and $p3$ is 2.756, 2.433 and 3.84 respectively. So, the order in which the pages will be represented is: at first $p3$, then $p1$ and finally $p2$.

IV. A K−NN Based Document Classification Algorithm

For the small domain *BUET*, we have proposed five predefined categories. They are: Undergraduate Admission, Graduate Admission, Course Info, Faculty Info and Others.

In the back-end of the system, each page is classified into the proposed five categories. Each *page-category* classification is weighted by some factor. The closer a page is to a category, the greater the weight value. At first, 60-70% pages were categorized manually and used as training data to categorize the rest of the pages. Distance weighted K−Nearest Neighbour [5] algorithm is then applied to all the uncategorized pages. As a result, each page is mapped to all the categories weighted by the *page-category* distance. Here $k = 10$ was used in our system and euclidean distance was used as a distance measure. In the front-end, the SERPs are grouped into two best categories and presented in a split screen format.

input : *TrainedDocs*[], a list of categorized pages
  TestingDocs[], a list of uncategorized pages
output: A set of newly categorized pages

C ← set of 5 predefined categories
foreach *page* TSP_i *in TestingDocs* **do**
  foreach *page* TRP_j *in TrainedDocs* **do**
    Calculate Euclidean Distance of TSP_i from TRP_j, $\mathrm{d}(\mathbf{TSP_i}, \mathbf{TRP_j})$
  end
  S ← set of 10 nearest training pages of TSP_i based on Euclidean Distance
  foreach *category* $cat_k \in C$ **do**

$$CS[TSP_i][cat_k] = \sum_{TRP \in S} \frac{1}{(\mathrm{d}(\mathbf{TSP_i}, \mathbf{TRP}))^2} \times \delta(cat_k, cat_{TRP})$$

    where $\delta(a, b) = 1$ if $a = b$ and $\delta(a, b) = 0$ otherwise
  end
end

Algorithm 2: Page-Category Weight Calculation

*1) **Training and Testing Phase**:* We have assigned a weight to each page-category mapping, enabling that page to be mapped to any of the predefined. First, we manually tagged 60% of the crawled pages to relevant predefined categories. One page may be tagged with multiple categories. The rest of the pages were then trained using k-nearest neighbor algorithm with $k = 10$. The system periodically crawls the entire *BUET* domain and categorizes any new uncategorized page in the same process. The process is described in Algorithm 2.

*2) **Searching and Ranking Phase:*** After assigning weight values to all the pages, for the predefined five categories, the preprocessing phase is finished. Now, we focus on the front-end algorithm; i.e. query evaluation and SERPs presentation. We assume the result window as a split window with two parts, each representing the pages of one of the two best categories that best suits with the user query. The algorithm for searching

input : $CS_{page,category}$, weight of a category for a given page.
userquery, a user query
output: A two category split screen display of SERPs

Apply Algorithm 1 for given *userquery*
Store the rank values in rank[*page*]
foreach *valid search result page* R_i **do**
  foreach *valid category* cat_k *of* R_i **do**

$$criteriaArray_{cat_k} \leftarrow criteriaArray_{cat_k} + rank[R_i] \times CS_{R_i,cat_k} \quad (4)$$

  end
end
Sort the *criteriaArray* on descending order of its values.
Let, cat_m and cat_n be the best two categories from the *criteriaArray* in terms of values.
Display the relevant SERPs of the best two categories in split window.
For each category, the result pages are ordered according to its rank value rank[R_i].

Algorithm 3: Query Evaluation and SERPs Categorization

and ranking phase first takes the user query as input and performs a traditional search using our proposed IR Scoring algorithm described in Section III. As a result, the pages are now ranked according to Algorithm 1. Then the pages are divided into two best categories according to Algorithm 3. The pages for the best two categories are then displayed in a structured format as if the display window has been split into two parts. The algorithm for query evaluation process in searching and ranking phase is given below:

TABLE IV. Weight of various categories for the SERPs obtained by traditional search

R_i	$rank[R_i]$	$CS[R_i][1]$	$CS[R_i][2]$	$CS[R_i][3]$	$CS[R_i][4]$
3	7.7	5.5	6.2	6.9	1.1
2	5.1	.75	1.125	8.9	5.3
4	3.9	1.8	3.3	6.6	7.1
1	3.6	4.4	2.9	5.7	9.3

*3) **A Sample Scenario:*** Lets assume four pages are found for a specific user query. Their rank values are calculated using the weighting scheme described in Section III. The calculation process for *criteriaArray* is given below:

$$criteriaArray[1] = 7.7 \times 5 + 5.1 \times .75 + 3.9 \times 1.8 + 3.6 \times 4.4 = 69.035$$

$$criteriaArray[2] = 7.7 \times 6.2 + 5.1 \times 1.125 + 3.9 \times 3.3 + 3.6 \times 2.9 = 76.78$$

$$criteriaArray[3] = 7.7 \times 6.9 + 5.1 \times 8.9 + 3.9 \times 6.6 + 3.6 \times 5.7 = 144.78$$

$$criteriaArray[4] = 7.7 \times 1.1 + 5.1 \times 5.3 + 3.9 \times 7.1 + 3.6 \times 9.3 = 96.67$$

Here, we can see that category 3 and category 4 are the best two categories for the given user query. Now, the system fetches each page according to descending order of the rank values ($R_3 > R_2 > R_4 > R_1$) and assigns its category from one of the best two categories that has the highest *CS* value for that page. For example, the system first fetches page 3. As category 3 has the higher *CS* value (6.9) than category 4 (1.1), page 3 is assigned to category 3. Finally, the display page is split into two parts to show the pages corresponding to these categories as follows:

TABLE V. Categorized SERPs

Category 3	Category 4
Page 3	Page 4
Page 2	Page 1

V. Experimentation and Results

The idea behind structured web search is to intelligently interpret user queries. Our structured prototype classifies search results into two categories that best match the given user query. In future, new categories may be introduced. Although the categories are predefined, they can be created dynamically in future using clustering techniques.

Traditional Web search in small domain concentrates on the domain's culture, users' search pattern and internet usage practises. The search results are presented keeping the traits of that domain in focus. In many cases, Google provides better results than *BuetSearch*. But in some cases, *BuetSearch* outperforms Google. For example, for the queries buet sports and sports, *BuetSearch* provides better result than Google as shown in Figure 2. For the query thesis cse, *BuetSearch* provides much better result than Google which shows web pages that were cited in other web sites and as a result has far better *PageRank*. On the other hand, *BuetSearch* highlights both graduate and undergraduate thesis results for CSE department. For the query cse course, results in *BuetSearch* is not so satisfactory. They are scattered and in some cases irrelevant. Google provides excellent result for this query. For the query undergraduate admission, *BuetSearch* succeeds to provide relevant results but the admission homepage does not appear as the first result. On the other hand, Google not only provides excellent results but present necessary links under the first search result. For space constraints, all the results could not be included.

Next, we shift our focus to the structured prototype. In Figure 3, with the query "course cse", the window

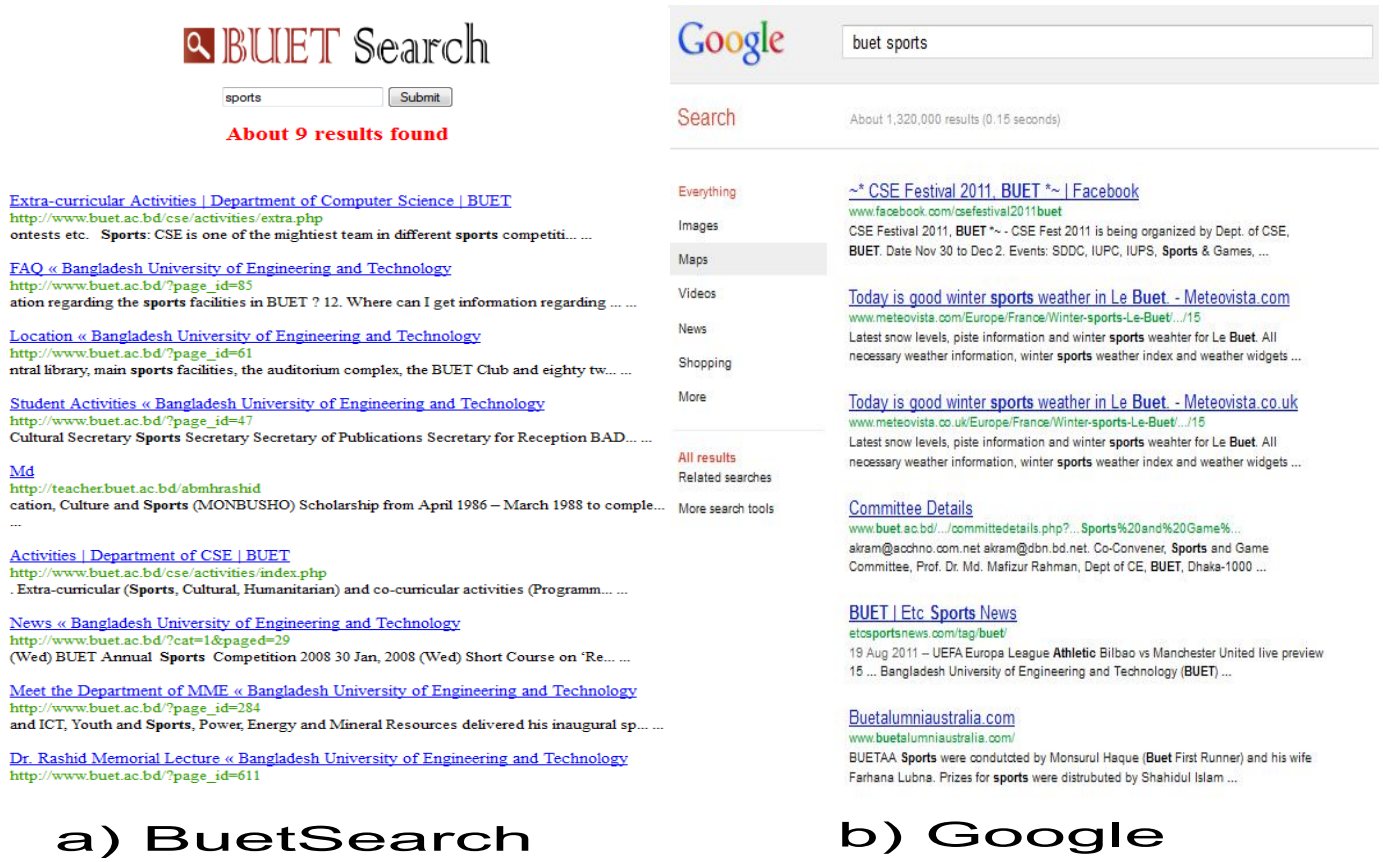


Fig. 2. Result representation in the structured prototype for the query "course cse"

is divided into a two category format: undergraduate admission and graduate admission. In this cases, the categorization is relevant. The corresponding SERPs for the categories are then shown according to their rank values. In the structured prototype, pages belonging to separate categories are ordered using the ranking and weighting schemes separately and presented in a two category format. With another query electrical faculty, the window is divided into a two category format: faculty info and course info. In the structured prototype, pages belonging to separate categories are ordered using the ranking and weighting schemes separately and presented in a two category format. In future, the number of categories can be extended to provide a more structured system. Table VI shows the precision of the structured prototype in categorizing SERPs for a given query. 20 such queries are shown here. The precision for the best and 2nd best categories are calculated separately. For each category c_i, the measure of precision is the percentage of SERPs that originally have category c_i as their best category among all the SERPs obtained for c_i. The analysis shows that, the structured prototype usually exhibits satisfying performance.

BuetSearch operates over the BUET domain and analyzes the query accordingly, whereas *Google* tries to interpret the query globally. Moreover, *PageRank* overpowers a lot of local pages of BUET domain. As a result, the SERPs provided by *Google* in this case are not satisfactory. So, web search in small domain can be a valid option and may outperform traditional large scale search engines in some cases. This concept can be extended for larger domains focusing on a specific community or a region.

TABLE VI. ANALYSIS OF THE PRECISION OF STRUCTURED SEARCH RESULTS

Query	Best Category Precision (%)	2nd Best Category Precision (%)
buet admission	69	97
walcom	100	100
cse department	66	100
eee department	73	75
admission	74	97
biis	72	96
civil department	66	100
cse400	50	100
course cse	79	64
notice	97	98
algorithm engineering	97	98
electrical faculty	71	63
kaykobad	72	100
undergraduate admission	85	33
mostofa akbar	100	100
artificial intelligence	100	80
course eee	36	100
student	88	33
exam	100	99
eunus	100	100

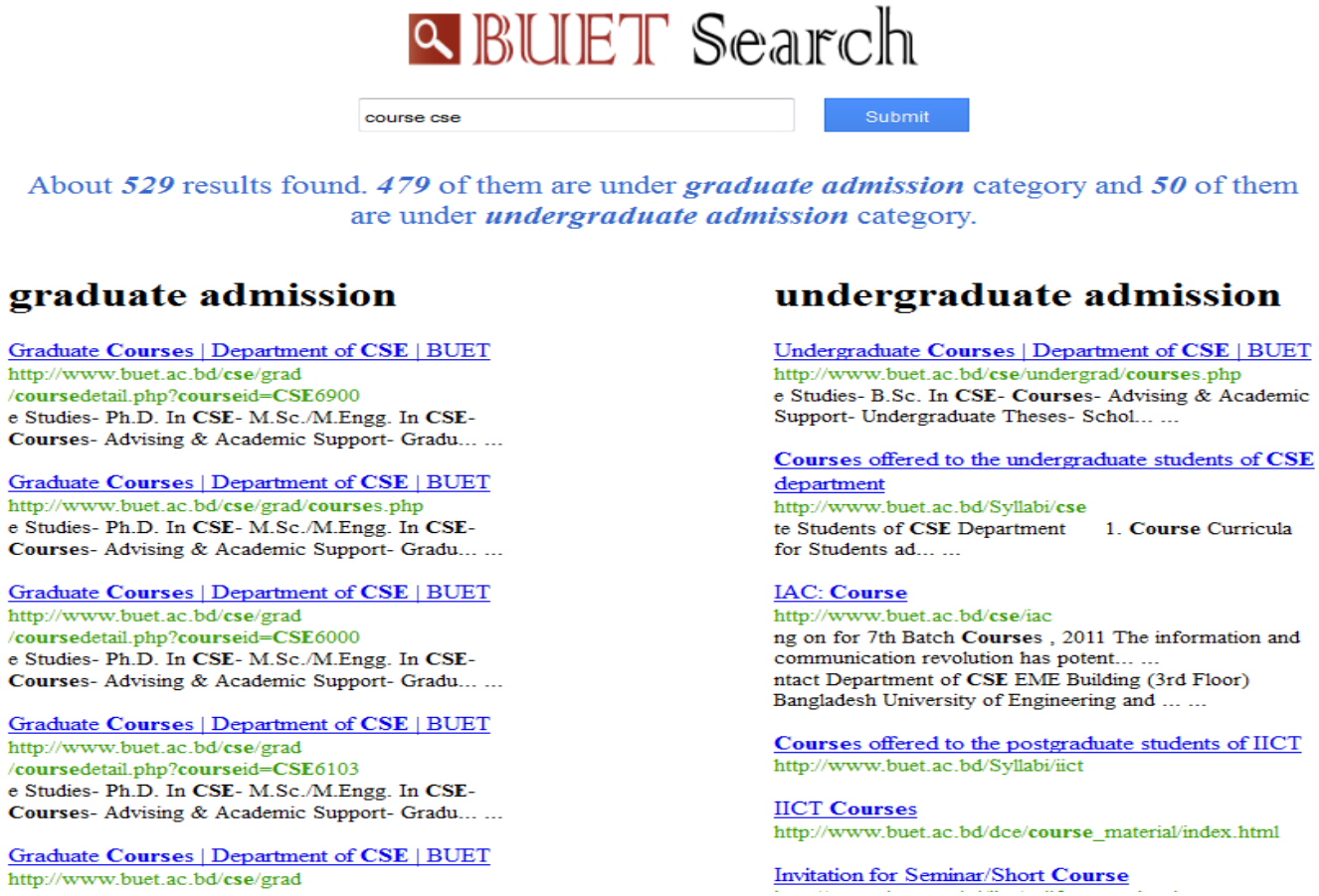


Fig. 3. Result representation in the structured prototype for the query "course cse"

VI. Conclusion

From the very beginning of the Internet, searching became an integral part of its users. With the increase in the number of sophisticated users, the complexity of search queries have increased. On the other hand, general users crave for better search results with simple queries. It is very difficult for a large scale traditional search engine to retrieve the required search result pages from a simple search query and satisfy the needs of all different and diverse people across the world who want different search results for the same query. To resolve this problem, our paper proposes the classic approach - divide and conquer; subdivide the web in scalable locations for simplicity and apply structured representation to deal with query complexity. Based on experimental results from *BuetSearch*, our paper has provided the idea of *Structured Web Search in Small Domain* as a probable effective solution. In future, new categories may be introduced. Although the categories are predefined, they can be created dynamically in future using clustering techniques.

References

[1] A. Aizawa. An information-theoretic perspective of tf-idf measures. *Information Processing & Management*, 39(1):45–65, January 2003.

[2] A. Blum, T. h. Hubert, C. Mugizi, and R. Rwebangira. A random-surfer web-graph model. In *ANALCO 06: Proceedings of the 3rd Workshop on Analytic Algorithmics and Combinatorics*, 2006.

[3] S. Brin and L. Page. The anatomy of a large-scale hypertextual web search engine. *Comput. Netw. ISDN Syst.*, 30(1-7):107–117, April 1998.

[4] P. Chebolu and P. Melsted. PageRank and the random surfer model. In *Proceedings of the nineteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 1010–1018. Society for Industrial and Applied Mathematics Philadelphia, PA, USA, 2008.

[5] S. A. Dudani. The distance-weighted k-nearest-neighbor rule. SMC-6:325–327, 1976.

[6] M. Hearst. Clustering versus faceted categories for information exploration. *Communications of the ACM*, 49(4):59–61, 2006.

[7] L. Page, S. Brin, R. Motwani, and T. Winograd. The pagerank citation ranking: Bringing order to the web. Technical Report 1999-66, Stanford InfoLab, 1999.

[8] G. M. Salton, A. K. C. Wong, and C.-S. Yang. A Vector Space Model for Automatic Indexing. *Communications of the ACM*, 18(11):613–620, November 1975.

[9] T. Seidl and H.-P. Kriegel. Optimal multi-step k-nearest neighbor search. In *SIGMOD Conference*, pages 154–165. ACM Press, 1998.

[10] M.-L. Zhang and Z.-H. Zhou. A k-nearest neighbor based algorithm for multi-label classification. In *Proceedings of the 1st IEEE International Conference on Granular Computing (GrC'05)*, pages 718–721, 2005.

A New Effective Part Selection Approach for Part-Based Gait Recognition

Md. Shariful Islam, Abdul Matin, Jyoti Paul,
Dept. of Computer Science and Engineering
Pabna University of Science and Technology
Pabna, Bangladesh
sharif_cse_ru@yahoo.com,
abdulmatin.cse090128@yahoo.com,
jyoti0790@yahoo.com

Md. Rokanujjaman and Md. Altab Hossain
Dept. of Computer Science and Engineering
University of Rajshahi
Rajshahi, Bangladesh
rokon_cstru@yahoo.com, altab_ru@yahoo.com

***Abstract*—In part-based human gait recognition, choosing the appropriate body parts is the most challenging problem. The various cofactors such as carrying conditions (backpack or side bag or hand bag), cloths (long coat or jacket or gown) affect various body parts. Here, we proposed a method for detecting the various cofactors in early stage. We have taken some pixel points which are marked as a boundary for each cofactor. The weight of these points is calculated and used to detect the particular cofactors in early stage. We divide the human body into seven body parts based on anatomical studies of gait. Discarding the body parts where the cofactor is present, the remaining parts are used in classification. We have achieved better result compared with other classical methods.**

***Index Terms*—Gait Recognition, Cofactors, Part-based.**

I. Introduction

Gait is a behavioural biometric that measures the way people walk [1]. The medical & physiological studies prove that human gait has 24 unique components [2] which are used as a great discrimination power for individual identification. When other biometrics such as face, iris and fingerprint need the subject cooperation but gait can be used without the subject cooperation at a distance [3]. Another advantage of gait is that it can be represented even at low resolution [1]. These distinct properties of gait make it applicable areas in video surveillance based like public transportations, banks, airports, and identifying perpetrators at a criminal case from a distance [4]. However, gait can be affected by various cofactors such as carrying bag (backpack, hand bag and side bag), wearing different cloths (jacket, long coat etc.).

There are mainly two approaches model based and appearance based. Model based approach [5] represents the gait using the body structure whereas the appearance based approach [6] represents the gait using the silhouettes in a gait sequence. Our work is appearance-based which is simple and fast. The most usable appearance based gait representation is Gait Energy Image (GEI) [7]. GEI is simply computed by taking the average of the silhouettes over a complete gait cycle. Example of some GEI images is given in Fig. 3 (first two columns and 4th column). GEI shows better result for the less intra-class variation (i.e., the subject does not carry an object or wear a complex type cloth). However it does not perform well when the intra-class variation is large. Recently, many part-based gait recognition approaches have been proposed to solve this problem [8], [9]. In [8], they divide the whole body into seven components (given in Fig. 1) and have reported that the combination of different body parts has increased the performance. In [9], they divided the whole body into eight parts and adaptively assigned more weight in unaffected body parts and less weight in affected body parts for clothing invariant gait recognition. The main problem of these part-based approaches is that to choose the unaffected body parts before matching measure. Another problem is that the combination of the individual part-to-part distance into a single distance in final classification.

In this work, we have proposed a comprehensive method for detecting the various cofactors in early stage. We have considered two sequences here one is normal and another is including a particular cofactor in training session. We get two average images for two sequences and it is used to define the set of points as a boundary for each cofactor. The weight of these points is calculated and it is used to detect the particular cofactor in testing session. We divide the human body into seven body parts. Excluding the body parts which are affected by the detecting cofactors, the remaining parts are used in final classification. To combine the individual part-to-part distances we have used the sum of products instead of summation because it can dominate the effective body parts in combinations.

II. Proposed Method

There are basically three cofactors: first one is simply normal clothing (i.e., subject does not carry an object or wear a different clothing combination), second one is carrying objects and last one is different types of coats. The carrying objects

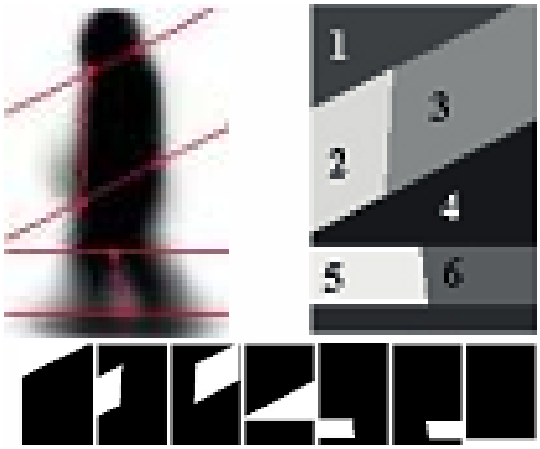


Fig. 1. Definition of seven body parts and corresponding templates [8].

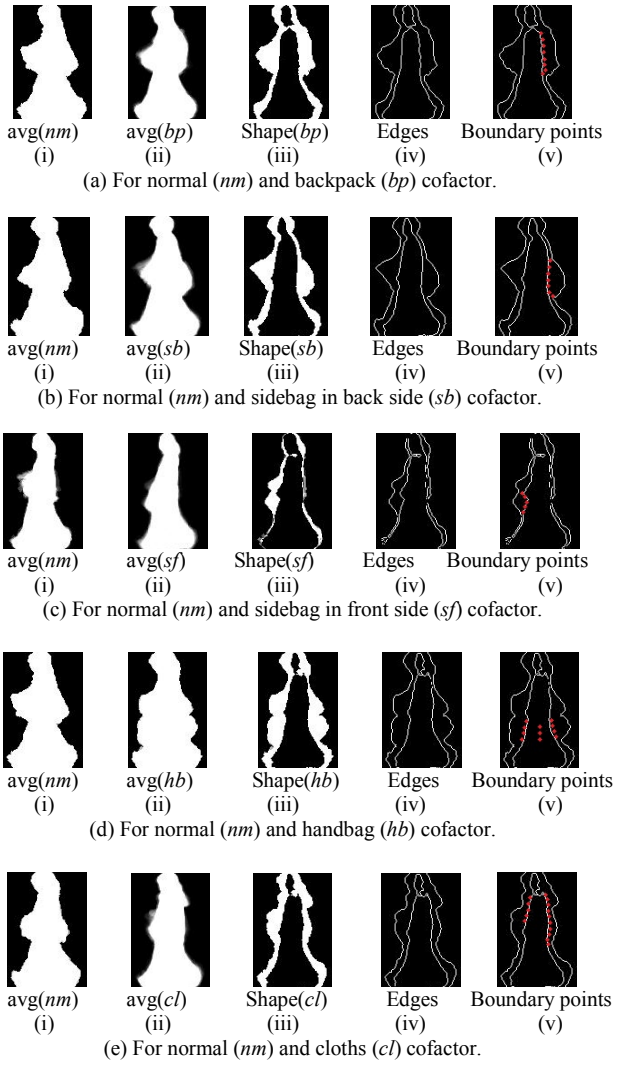


Fig. 2. Images to setup the boundary points for each cofactor: (i) Average images of all normal (*nm*) sequences (ii) Average images of a particular cofactor (*cf*) (iii) Difference images between the *nm* and *cf* (iv) Images of edge in (iii) and (v) Images of the set of boundary points for each *cf*.

are further divided into four sub cofactors:backpack, side bag at front side in arms swing area, side bag at back side in arms swing areas and handbag. We have defined here a set of points that are used as a boundary values for each cofactor. These set of boundary points are used to detect the particular cofactor in testing session. We have used the most using gait representation GEI as an input gait features. For each cofactor we setup a training set considering 10 different subjects randomly out of all subjects with this cofactor in the CASIA dataset B [12]. Each subject in the training set for each cofactor contains two sequences one is normal and another is the targeting cofactor respectively.

First we compute the two average GEIs using all the sequences in the training set of each cofactor: one for normal (*nm*) sequences and another for the sequences with the targeting cofactor (*cf*), {*cf* = *clothing (cl), backpack (bp), sidebag at back side (sb), sidebag at front side (sf), and handbag (hb)*}.

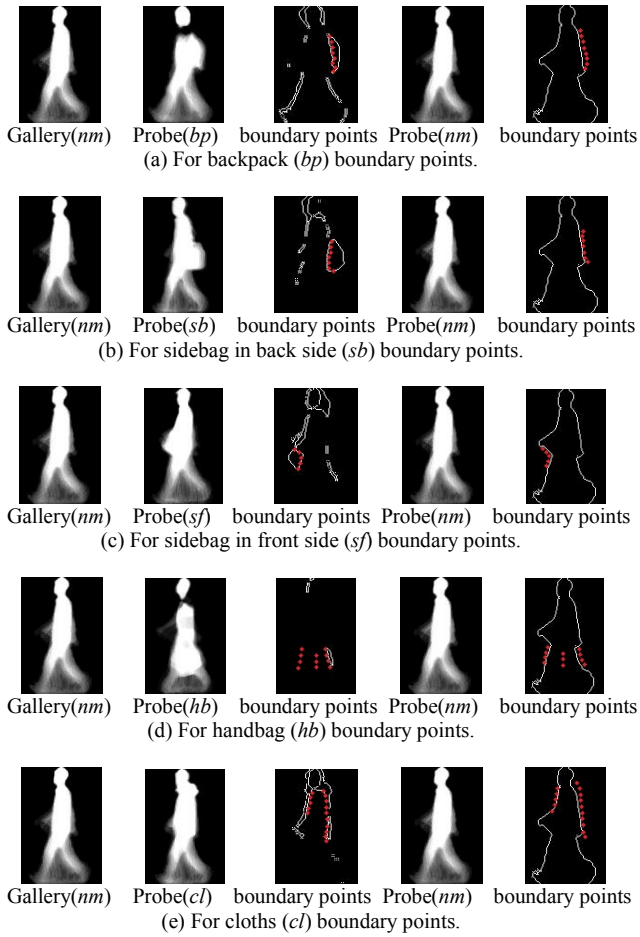


Fig. 3. Cofactor detection using the set of predefined boundary points for a new probe subject with either normal (*nm*) or particular *cf* cofactors.

Let there are *m nm* GEIs and *n* targeting *cf* GEIs in the *cf* cofactor training set. The average GEIs for *nm* sequences is computed as:

$$Avg_{nm}^{cf} = {}^{1}/_{m} \sum_{i=1}^{m} GEI_{nm}^{cf}(i) \qquad (1)$$

where, $GEI_{nm}^{cf}(i)$ is the ith *nm* GEI in the *cf* cofactor training set respectively. Similarly, the average GEIs for a cofactor *cf* is computed as:

$$Avg_{cf} = {}^{1}/_{n} \sum_{i=1}^{n} GEI_{cf}(i) \qquad (2)$$

where, $GEI_{cf}(i)$ is the ith *cf* cofactor GEI in the *cf* cofactor training set respectively. Now we have two average GEIs: one for normal Avg_{nm}^{cf} and another for a *cf* cofactor Avg_{cf} for each *cf* given in Fig. 2 (i) and (ii). Next we compute the shape or outlier of each *cf* cofactor by subtracting Avg_{nm}^{cf} from Avg_{cf} as:

$$Shape_{cf} = \sum_{i=1}^{r} \sum_{j=1}^{c} Avg_{cf}(i,j) - Avg_{nm}^{cf}(i,j) \qquad (3)$$

The computed $Shape_{cf}$ of each *cf* cofactor is given in Fig. 2 (iii). Now we find out the edges of the computed $Shape_{cf}$ of each *cf* cofactor using the Sobel Edge Detector algorithm. The images of the edges of each cofactor *cf* are given in Fig. 2 (iv). Finally, we manually select the set of inner points on the edge that are fit into the targeting cofactor *cf* as the boundary values given in Fig. 2 (v). These boundary values are used to detect the targeting cofactor in testing session.

A. Cofactor Detection Process

Let there are k points for a *cf* cofactor that are fit as boundary values. Then the sum of intensity values of Avg_{cf} for these k points are used as the weight (W_{cf}) for the targeting *cf* cofactor. Therefore, there are k predefined points for each *cf* cofactor and each cofactor has a weight (W_{cf}). To test the existence of a cofactor *cf* in a new query probe sequence, we compute the weight (W_p) that is the sum of intensity values of the probe subject for the predefined k points. If the calculated weight W_p is less than the predefined weight W_{cf} for the *cf* cofactor then the subject is safe that is the *cf* cofactor does not present here otherwise the *cf*cofactor is present. For example, the cofactor detection process is given in Fig. 3 for a gallery (normal standard cloths; first column) and a probe with *cf* cofactors (2nd column) or normal (4th column). We can see that our predefined set of boundary points clearly detect the targeting *cf* cofactor when it is present in probe (3rd column) and also detect as outside when it is not present in probe (last column).

B. Body Parts Definition and Effective Part Selection

We divide the human body into seven body parts (Head+Neck, Chest, Abdomen, Pelvis, Thigh, Shank, and Foot) based on anatomical studies of gait [10] given in Fig. 4. When a cofactor is detected then the corresponding combination of body parts are discarded in early stage. The remaining body parts are selected as effective parts and used in final classification. The selected effective body parts and discarded affected body parts for each *cf* cofactor are given in Table I.

III. Classification

Each gait sequence consists of many gait cycles and each cycle is represented as GEI here. Assume there are c classes of gait sequences G^k $\{k=1,...,c\}$ in the gallery set. Let a probe sequence P with s GEIs $P_i\{i=1,..,s\}$ and a gallery sequence G^k with t GEIs G_j^k $\{j=1,...,t\}$. The matching measure for the GEIs is simply chosen as the Euclidean distance between P_i and G_j^k (let, $d^{ij}(P_i, G_j^k)$). We have taken the minimum of the minimum distances of the GEIs as it is effective when some gait cycles or GEIs contains more noises than other GEIs in a gait sequence. The minimum value of the minimum distances of the combinations of GEIs of each probe and gallery is defined as:

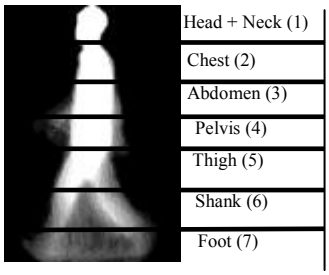


Fig. 4. Seven body parts definition

TABLE I. Selecting And Excluding Body Parts When A Cofactor is Detected.

Cofactors	Selecting parts	Excluding parts
Backpack	**1,4,5,6, and 7**	2 and 3
Side bag	**1,2,5,6, and 7**	3 and 4
Handbag	**1,2,3,4, and 7**	5 and 6
Cloths	**1,5,6 and 7**	2,3, and 4

$$d^{Min}(P, G^k) = Min_i[\underset{j}{Min}\{d^{ij}(P_i, G_j^k)\}] \quad (4)$$

In part-based classification, the main problem is how to combine the individual part-to-part distance into a single distance in final classification. In this case, we have used the sum of products instead of summation to combine all part-to-part distances into a single distance. Because it can dominate the major body parts which are more effective among all the several body parts. So the overall matching measure between probe and gallery is:

$$D(P, G^k: G^*) = \prod_{i=1}^{N} d_i(P, G^k: G^*) \quad (5)$$

Where, N is the number of body parts and $d_i(P, G^k: G^*)$ is the distance between P and G^k for the ith body part and G^* is the set of sequences of all the galleries.

IV. Experiments

A. CASIA Dataset

The CASIA gait database [12] is used in this work. It consists of 124 subjects. For each subject, there are 10 walking sequences: six normal walking sequences where the subject does not carry a bag or wear a coat, two carrying bag (backpack, hand bag, side bag.) sequences and two wearing different cloths (jacket, long coat etc.) sequences. We have used the GEI as an input gait features. Examples of GEI images for different cofactors used in this dataset are given in Fig. 3 (first two columns and 4th column). The image size of each GEI is 128×88. The gallery set is constructed using the first four sequences from the six normal sequences. The probe set is constructed using the rest of the two normal sequences, two carrying bag sequences and two wearing different cloths sequences.

B. Comparison Algorithms

We compare the proposed method with the three existing methods [11], [7], [8]. In [11], they proposed a baseline algorithm or direct template matching. In [7], they apply the Linear Discriminant Analysis (LDA) on the dimension

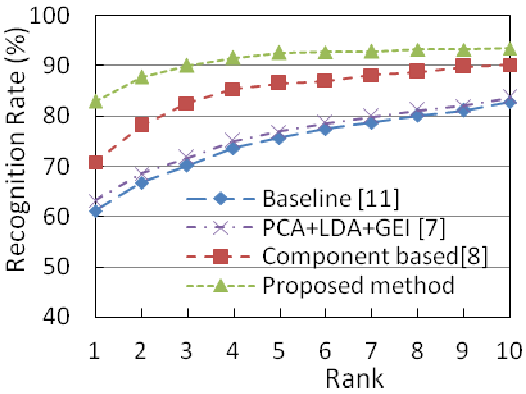


Fig. 5. Algorithm comparisons

reduced Principle Component Analysis (PCA) space to get the best data representation and class separability. In [8], they divide the human body into seven components and the combinations of these components are used for classification.

C. Result and Discussion

The performances of all the methods are demonstrated in Fig. 5. It is observed from the Fig. 5, the proposed method shows much better result in a wide rank wise evaluation. We discard the affected body parts using our defined set of boundary points for each cofactor dynamically in testing session and get the better result. Again in part-based matching measure, the sum of products of the individual part-to-part distances (using Eq. 5) can dominate the most effective body parts in final classification than the straight forward summation of each part-to-part distances. We can see that from the Table II, the sum of products always shows better performance than the summation based combination techniques for each cofactor in part-based classification. We summarize the detection rates of each cofactor in Table III. Here we consider a number of sequences of each cofactor to find the effectiveness of the defined set of boundary points for each cofactor. We can see that the detection rates are in acceptable for all cofactors. It is observed that, all the backpacks are easily detected by the set of boundary points. It is also still better for clothing cofactor. The detection rates are comparatively low for handbag cofactor because the changing areas of gait by this cofactor are overlapped with the legs swing areas of gait. The previous works show very poor performance especially for clothing cofactor but the proposed method is still shows better performance given in Table II.

VI. Conclusions

This paper proposed a new effective part selection approach for part-based gait recognition. It is defined a set of points as a boundary for each cofactor. The weight of these points is used to detect the presence of a particular cofactor. We divide the human body into seven parts. Discarding the affected body parts where the cofactor is present, the remaining body parts are selected as the most effective parts in final classification. The proposed method shows much better result when it is compared with the other classical methods. In future, we will try to develop a curve fitting algorithm that cope the cofactor as an outlier object for more robust gait recognition.

TABLE II. Performances For The Combination of Part-to-Part Distances Into A Single Distance

	Combination techniques	
Cofactors	***Summation (%)***	***Sum of products (%)***
Normal	100	**100**
Carrying Objects	84.27	**85.08**
Cloths	45.96	**63.71**

TABLE III. Performances of Cofactor Detection

Cofactors	No. Of Sequences	Detected sequences	Percentage
Normal	535	492	91.96
Backpack	33	33	100
Side bag (back side)	128	99	77.34
Side bag (front side)	17	12	70.58
Handbag	36	24	66.66
Clothing	214	188	87.85

References

[1] L. Lee, W. Grimson, Gait analysis for recognition and classification, in: Proceedings of the Fifth IEEE Conference on Face and Gesture Recognition, vol. 1, 2002, pp. 155–161.

[2] M.P. Murry "Gait as a Total Pattern of Movement," American Journal of Physical Medicine, vol. 46, pp. 290-332,1967.

[3] K. Bashir, T. Xiang and S Gong "Gait recognition without subject cooperation", Pattern Recognition Letters vol. 31, pp. 2052-2060, Oct. 2010.

[4] D. Hatzinakos and K. Plataniotis, "Gait Recognition", Bell Canada Chair in Multimedia. IPSI: Identity, Privacy and Security initiative @ UofT, The Edward S. Rogers Sr. Dept. of Electr.&Comput. Eng, University of Toronto, M5S 3G4, Canada.

[5] Yam, C., Nixon, M.S., and Carter, J.N.: Automated Person Recognition by Walking and Running via Model-based Approaches, Pattern Recognition, Vol. 37, No. 5, (2004) 1057–1072

[6] L. Wang, T. Tan, H. Ning, W. Hu, Silhouette Analysis-Based Gait Recognition for Human Identification, IEEE Trans. on PAMI Vol.25, No. 12, Dec.,2003.

[7] J. Han and B. Bhanu, "Individual recognition using gait energy image", IEEE Trans. PAMI vol. 28, pp.316-322, 2006.

[8] X. Li, S. Maybank, S. Yan, D. Tao, D. Xu, Gait components and their application to gender recognition, IEEE Transactions on Systems, Man, and Cybernetics Part C 38 (2) (2008) 145–155.

[9] M.A. Hossain, Y. Makihara, J. Wang, and Y. Yagi, "Clothing-invariant gait identification using part-based clothing categorization and adaptive weight control", Pattern Recognition, vol. 43, pp.2281-2291, 2010.

[10] W. Dempster, G. Gaughran, "Properties of body segments based on size and weight", American Journal of Anatomy 120 (1) (1967) 33–54.

[11] Z. Liu, and S. Sarkar, "Simplest representation yet for gait recognition: averaged silhouette," International Conference on Pattern Recognition (ICPR), vol.4, pp. 211- 214, 2004.

[12] S.Yu, D. Tan, and T.Tan. "A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition," In Proc. Of the 18th Int. Conf. on Pattern Recognition, vol. 4, pp. 441-444, 2006.

A Switching Mechanism for Grid Providers

Aminul Haque
Department of Computer Science & Engineering
Daffodil International University
Dhaka, Bangladesh
aminul.cse@daffodilvarsity.edu.bd

***Abstract*— Economic-based resource management models are found to be efficient for distributed resource collaboration in the Grid. Due to the wide range of applications and dynamic nature of Grid resources, constant performance by a specific economic model is therefore subjective. This work identifies the potential of different models in different scenarios in the Grid. This identification motivates us to design an optimization framework that couples suitable economic models and is able to switch from one model to another based on the models' domains of strength. The research further describes the roles played by an agent in dynamically deciding which model to be used when and for what purpose. It shows the effectiveness of the switching framework compared to any other individual models in a dynamic computing environment.**

***Index Terms*— Grid, economic models, domains of strengths, switching, agent, dynamic**

I. Introduction

The major motivation behind the Grid is to solve computationally complex problems such as drug design in a standard and cost effective way. The Grid can be distinguished from conventional computing platforms due to its distributive, dynamic, and heterogeneous (independent) nature of resources. At the same time, this various dimension of resources creates a challenge to seamless resource collaboration which is essential for building a large scale autonomous computing platform. Economic-based approaches are found efficient to deal with this challenge as identified by many researchers [1-3].

Economic models are efficient for Grid computing as they can be used to control and regulate the behavior of Grid entities by providing necessary standards and tools. In addition, economic approaches provide sufficient motivation to the resource providers to contribute their resources in the Grid. In spite of potential of the economic-based management, different economic models are based on different working principles and pricing methodologies.

It is often impossible to claim the suitability of a particular model in every aspect of a highly dynamic and distributed environment such as Grid. In addition, the domain of Grid applications has been broadening day by day where the scheduling policies related to the applications and objective functions regulated by resource providers are hard to be justified by a single model. A survey paper on existing economic models also concurs with the fact of the suitability of different models in different scenarios [4]. For example, English Auction (EA) economic model has been found suitable for maximizing revenue for providers; however, the model generates a higher communication cost whereas Commodity Market Model (CMM) is found suitable for maintaining equilibrium between supply and demand of resources.

The survey work shows the opportunity of developing an optimization framework that couples suitable economic models and is able to switch from one model to another based on the models' domains of strengths. However, to understand the domains of strength of individual models, a quantitative analysis on the performances of the models is important. In [5], such analysis has been carried out. Based on this analysis, domains of strengths and weaknesses of individual models for a comprehensive set of scenarios have been identified. Being motivated from this identification, this research develops an optimization framework through switching from one model to another in a dynamic computing environment. The major contributions/challenges in developing the switching model of this paper are as below:

- The design and development of a switching framework which is capable of adaptive resource management procedures and optimizing various performance metrics as predefined by Grid resource providers
- The development of an agent designed for dynamically providing decision to the framework on which economic model to be used when and for what purpose

The rest of the paper is organized as follows. In Section 2, related works and the motivation for our work are provided. Section 3 gives a brief overview of widely discussed economic models in the Grid. A short discussion on identifying the domains of strengths is presented in Section 4. Section 5 incorporates the design and development of the switching framework and the switching agent. Section 6 discusses the evaluation of the framework.

II. Related Work

As described above, suitability of different economic models for different scenarios was established in [4]. In [5], domains of strengths and weaknesses of these models were identified. This section briefly discusses the findings of these papers and states the focus of this paper.

Richard et al. propose that Commodity Market Model would be suitable for maintaining market equilibrium and minimizing communication cost compared to the English Auction [6]. On the other hand, Tan and Gurd criticize Commodity Market Model due to its system-oriented approach rather than being incentive-oriented [7]. They argue, in

Commodity Market Model, price formation process considering global information on supply and demand does not account for individual's preference optimization; thus become undesirable for the participants. Continuous Double Auction (CDA) is proposed to be suitable compared to single-sided auctions such as English Auction in terms of communication and time efficiency.

It has been identified that English Auction is suitable to maximize revenue, economic efficiency (Pareto-optimality in resource allocation) and the QoS [2, 3]. Contract Net Protocol (CNP) has been found to be suitable for the utility-based resource allocation and scalability [8]. It is also suitable for solving distributed cooperation problem and for meta-scheduling process. On the other hand, in a distributed environment such as Grid, Bargaining Model (BAR) is proposed to be suitable; because it supports utility-based negotiation between a user and a resource provider [9].

Realizing this variation in the performances of different models, a quantitative analysis has been conducted to measure and contrast the effectiveness among the various widely proposed economic models in Grid resource management [5]. The paper's experimental model and resource configuration help identify the domains where a particular model outperforms the other models. The paper shows promising outcomes in supporting the methodology of utilizing the potential of different models in different scenarios. These findings work as a motivation for developing an optimization framework which incorporates adaptive resource management capabilities while switching from one model to another. A detail analysis on the domains of strengths of individual models can be found in [5]. This paper only briefly describes the identification of the domains here. The focus of this paper is therefore to elaborate the discussion on the development of the optimization framework using the five most widely proposed economic models in the Grid.

III. Widely Proposed Economic Models in the Grid

Economic models are significant for Grid resource management; because they can explain the behavior of Grid entities and predict the impacts of the models on the market. Therefore, it is important to understand the parameters that define the models. Please refer to [5] for a detailed discussion on the implementation of widely proposed economic models in the Grid. This paper only presents a brief overview of the models here.

Commodity Market Model. There are two types of CMM in the literature; one is flat pricing in which price for a resource does not vary frequently and another is supply and demand driven in which the price changes very frequently based on the market's supply and demand function. The latter is widely adopted and more popular in the Grid [10]. Thus, this work considers the supply and demand driven CMM. The essence of this type is to change the resource cost frequently based on current supply and demand function. The price that brings equilibrium between supply and demand in the market is regarded as equilibrium/spot price.

Bargaining Model. BAR is one of the most widely proposed economic models in the Grid [4]. It supports negotiation between users and providers over multiple rounds. Such negotiation eventually helps to better understand the requirements of the market entities. The negotiation facilitates Service Level Agreement (SLA) processes and brings Pareto-optimality in resource allocation for users. Here, a user might start with very low bid and a resource-node with very high bid and the negotiation process continues until they reach to a mutually agreeable condition or any of them does not show any interest to continue [1].

English Auction. It has been identified by many researchers that auctions are suitable for distributed resource collaboration among autonomous and self-interested entities [11, 12]. EA of its ascending bid type is the most popular in the Grid [4]. The auction is typically conducted by an auctioneer who keeps increasing the bid over each round. Once the total number of rounds finishes or no broker is willing to accept the current bid, the auctioneer checks its reservation price and selects a winner. In case of a failure, the broker can again participate in another auction and continue the competition.

Continuous Double Auction. The CDA of its most popular form which is open cry with order queue has been implemented in this work [7]. In this form, resource costs are generated continuously until the Auctioneer finds a match between a bid (broker request) and an ask (resource properties). Bids and asks can be placed any time during the auction phase. Outstanding bids and asks are maintained in an Order Book while bids are sorted in descending order and asks are in ascending order. Whenever there is a match, the auctioneer immediately informs the message to the corresponding resource so that it can stop generating asks further. The auctioneer also sends a message to the respective broker so that it can submit its Gridlet to the resource.

Contract Net Protocol. CNP is popular in Grid computing, especially because of its support for meta-scheduling and for resource cooperation to deliver a strong QoS to the users. A broker and a resource-node are known as manager and contractor respectively in this scenario [1]. In such a market scenario, a manager tries to optimize its scheduling process (meta-scheduling) by selecting one or more suitable contractors from available contractors in the market. The selection process is typically conducted according to the manager's preference values. In terms of a task execution in a distributed environment such as Grid, the manager could either prefer time or budget to be optimized.

IV. Domains of Strengths of Different Economic Models in the Grid

Economic models have the ability to serve the Grid from various dimensions. It helps understand a user's QoS requirements and resource values to the providers. Therefore, understanding the strengths and weaknesses of a particular model for specific scenarios is crucial. As mentioned earlier, a detailed analysis on domains of strengths have already been given in [5]. The analysis has been conducted for at least 10 performance metrics.

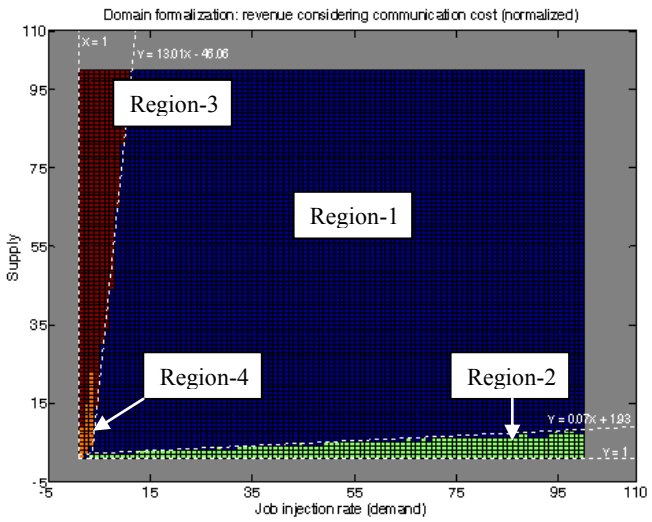


Fig. 1. Domain analysis for revenue over communication cost

This work presents the discussion only on revenue considering communication cost (overhead) metric for concision.

Revenue Analysis. The contour diagram (Fig. 1.) shows the result. One can observe from the contour that the EA is fully absent here rather CDA, BAR, CNP and CMM are the models show their strengths at different scenarios. Out of the four models, CDA shows the best performance. The maximum part of the simulation space (Region-1) is dominated by CDA. There are two different reasons for showing such performance by CDA. Firstly, due to the sorting process in CDA, it generates the lowest communication overhead in most of the scenarios. This prevents the model from lowering the revenue much. Secondly, due to the special price formation process *(askPrice+bidPrice)/2* of the model, the revenue earned from a particular trade lies somewhere in between the Gridlet's budget *(bidPrice)* and the resource's *cpu-Cost (askPrice)* [5]. Therefore, the ultimate revenue is neither very low nor very high. However, when the demand is sufficiently low irrespective of the supply (Region-3), BAR outperforms all the others. Due to the high supply, Gridlets get high chance to start bargaining with appropriate resources. Thus, the Gridlets make quick acceptance on the resources, which leads to a lower communication overhead for BAR. In addition, due to the low demand and high supply, Gridlets do not need to switch between the resources very often if a resource is busy communicating with others.

When both supply and demand are low (Region-4), CMM tends to outperform other models. However, due to the little contribution, one can ignore this region and assume that this is a part of Region-3. As the supply is sufficiently low regardless of the demand (Region-2), CNP performs better than the others. CNP, in this region, produces lower communication overhead compared to that of the CDA. Moreover, as CNP supports both time and cost optimization for the Gridlets, the decreased revenue during budget optimization could somehow be covered by the increased revenue during time optimization scenarios. Therefore, like CDA, CNP produces a kind of intermediately revenue overall.

One can observe that a single model does not perform better all the time. This proves the compatibility of the findings with existing literature, i.e., in a highly dynamic and distributed environment, a single model is not suitable to cope with every scenario [4]. From this identification, this research finds the opportunity of optimizing different performance metrics by utilizing the potential of different models in different Grid scenarios. The following section presents a framework suitable to process and evaluate such optimization.

V. An Optimization Framework for Economic Grids

The following example provides some motivation for optimization using multiple economic models in the Grid.

Example: Let us imagine a Grid network that would like to maximize its revenue and minimize communication overhead. This work has identified that there are three different regions where three economic models perform better compared to each other in terms of revenue over communication cost objective function (Fig. 1.). Now, the possibility of maximizing the revenue is higher if the network employs the three models in their respective domains rather than keep rely on a single model throughout the whole space.

For example, if the network notices that its current demand is low regardless of its supply, it can employ BAR as BAR has been identified as the suitable model for this scenario (Fig. 1.). Likewise, if sometimes the network notices that its supply has been decreased moderately and demand has been increased, it can switch to CDA. Again, when the supply becomes sufficiently low and demand is still high, the network can switch to CNP. As a result, the network would be able to utilize the potential of different economic models in different scenarios, thus maximizing the defined objective function in general.

Similar optimization is also possible for other metrics where different models perform in different scenarios. Such optimization procedure would provide incentives to Grid providers which in turn would help to deliver a viable market mechanism for future computing platform. However, to be consistent with the nature of the Grid, such a multi-model architecture gives rise to the following questions.

- What motivates a Grid to follow such an optimization procedure?
- How to utilize the potential of different economic models in a highly dynamic environment?
- Who is responsible to keep track of the scenarios?
- Who is responsible to decide which model to use and when to use?
- Who is responsible for the corresponding effects of changing models?

To deliver the answers of these questions, this work proposes a switching model that dynamically switches from one economic model to another and is able to adapt with the changing circumstances in the environment. To facilitate the switching process, the work designs a switching agent, which automatically decides which model to be used when and for what purpose. A prototype of our switching model is presented in the following section.

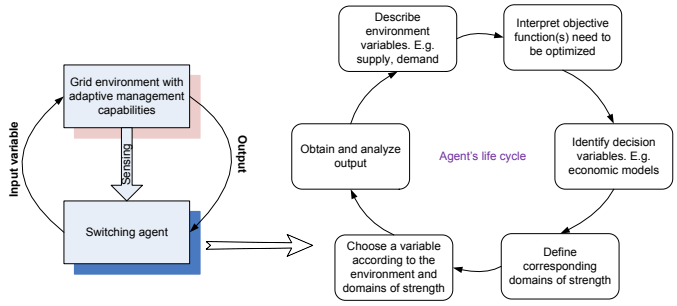


Fig. 2. Agent-driven Switching Framework for Grid Resource Management

A. A Prototype of Switching Framework

To understand the working behavior of switching framework, a prototype that describes the key functionalities performed by the framework is designed in Fig. 2. .

Two different parts of the framework will be described here; the Grid environment with adaptive management capabilities and the switching agent. The roles (broker, resource) in such an environment must have capabilities to deal with different models dynamically. Therefore, the parameters and organization of a broker/resource model have to be designed in a way so that it can adapt with changing behavior of the environment. The auctioneers in the auction models, in this case, must be prepared so that they can start processing as soon as any one of them is invoked by the system. In addition, every broker, resource and auctioneer must understand the language of the agent; because the agent is providing crucial information to the system. For example, the agent informs the participants in the environment about whether to continue with the current model or to get prepare for a new model. Therefore, the ability by the participants to interpret the agent's message and to work accordingly is significant.

The most crucial role played in the environment is the agent. There are six different stages in the agent's life-cycle. The agent can sense (keep track of supply and demand in the network) the Grid environment in two different ways; time-based and event-based. In a time-based mode, the agent senses the environment at every specific period of time whereas in an event-based approach, the agent senses the environment whenever there is a change in supply and demand in the environment. This research considers the event-based approach and would like to explore the time-based scenario in future.

Whenever there is a change in supply or demand, the agent is invoked to provide decision on whether to continue with the existing model or to switch to another model. The agent, at first, interprets the current supply and demand, and the objective function that needs to be optimized. Based on the objective function, the agent then identifies the available economic models those can help optimize the defined objective function, where the models are considered as decision variables. For example, in terms of revenue considering communication cost, there are three different decision variables; BAR, CDA and CNP (Fig. 1.). Each of these models shows the strength over the others for its respective regions. The following process by the agent is to identify the corresponding domains of strength of the decision variables.

Pseudocode-1: Adaptive Broker Model

```
1. Switch (Track messages using Tag values)
2.  {
3.    Case: Msg. Received from Environment//scheduled by Pseudocode-2
4.          Cast the msg. and identify the Recommended_model
5.          if (Current model != Recommended_model)
6.          {
7.             Stop sending request from this broker
8.             Wait for a while //this is important if the broker is
                        //currently busy negotiating other resources
9.             De-activate the parameters related to Current model
10.            if(The broker is still looking for suitable resources)
11.               {let delay = 0;
12.                Current model = Recommended_model
13.                Send a msg. to itself to prepare for the new model
14.                sendMsg(this.Broker, delay, Tag.Current model);
15.               }else terminate the broker
16.         }else keep continue with the Current model
17.   Break;
18.
19.   Case: Request Send
20.         Prepare the job specification according to the model
21.         Start sending the request to the resources one by one
22.   Break;
23. }
```

Depending on the values of current supply and demand, the agent verifies the appropriate domain and decides the model corresponding to that domain. If the decided model is different from the previous model, the new model works as an input to the environment (Fig. 2.). Otherwise, the system continues with the existing model. The process continues until the supply or demand gets unavailable from the environment. The following section describes the development process of our switching framework.

B. Framework Development

The rising demand of Grid computing has led to the requirement of a framework that support dynamic organization of different models. The user (broker), in this case, must define the parameters required to deal with all the economic models in the framework. This helps the corresponding broker to be adaptive with the environment where the interaction principles may change over times. Apart from some common parameters, the user must specify the parameters more particular to a specific model.

The broker model suitable for the switching framework is described in *Pseudocode-1*. There are two major parts in this code. The first part deals with the interpretation of the message received from the environment about the recommended model and subsequent procedures to get ready for the new model (line3 through line17). The second section of the pseudo code corresponds to the construction of the query according to the new model (line19 through line22). The *Pseudocode-1* is in action until the simulation finishes or the broker is terminated. A similar model is implemented for resources. The characteristics of our framework when it switches from one model to another are discussed below.

Autonomous Switching. The rising complexity in computing systems has led to the challenge for a feasible understanding about the future of those systems to deal with dynamic and conflicting demands in the systems. An autonomic computing system is typically defined as a system that can evolve by self-management process in order to meet the system-administrator's goal. This sub-section discusses about such an autonomous system that is able to deal with dynamic integration of different economic models in the Grid.

Pseudocode-2: Environment Management – Broadcast Switching Decision – Switch from model-1 to model-2

```
1.   Msg. Received from Agent  //scheduled by Pseudocode-3
2.   The agent suggests model-2 for current network scenario
3.   Stop initiating auction instances if the model-1 is auction
4.   Wait for a while to complete model-1 executions
5.   De-activate the global parameters related to model-1
6.   let delay = 0;
7.  //Send a msg. about model-2 to every broker in the network
8.  Foreach(broker in the broker-List)
9.  {
10.    Send a msg. about model-2 to the broker without any delay
11.    sendMsg(this.Broker, delay, Tag.Recommended_model);
12.  }
13. //Send a msg. about model-2 to every resources in the network
14. Foreach(resource in the resource-List)
15.  {
16.    Send a msg. about model-2 to the resource without any delay
17.    sendMsg(this.Resource, delay, Tag.Recommended_model);
18.  }
```

Upon receiving the message of a new model from the agent, the environment starts processing its subsequent procedures. The procedures include dealing with parameters to control the behavior of the current model and to inform the participants in the environment about the new model. *Pseudocode-2* describes these contributions in detail. The environment is the first entity that receives the notification of switching. The entity has been constructed with general components (e.g. broker-list, resource-list), managing auction entities (e.g. creating auction instances, recording winner details etc.), and informing network entities about the new model before the model starts its action. Once the environment notices that the agent (line2) has suggested a new model, it waits for a while to neutralize the effect of existing model (line4) and quickly broadcast the decision to existing brokers and resources (line8 through line18) through the network.

The main parameter which drives the switching framework is the optimization function by the Grid provider. The switching agent must be informed about this function in order to perform the decision process properly before the simulation starts. The following section describes the functions implemented by the switching agent.

C. Switching Agent

"Learning" – is an important term in the field of agent technology. It is now a closely related topic between natural and computational systems. It is typically used to mean the improvement of a system based on past experience. The objective of autonomous system and learning is closely coupled. However, leaning is particular in dealing with individuals. This work develops an agent that is able to apply its knowledge based on previous identification (domains of strengths) towards improving the performance of the system.

The agent now needs to apply this knowledge in a dynamic Grid environment and to prove its effectiveness in terms of optimizing different performance metrics. An example of the agent's knowledge representation is presented below.

The Agent's Knowledge Representation: In our framework, the agent uses the economic models as input variables for optimization in the environment. As this work deals with five models, the problem space defined by the agent is,

Problem Space = {CMM, BAR, EA, CDA, CNP}

Pseudocode-3: Switching Decision

```
1.   Input: Demand,Supply,Optimization-metric,current_model,gradient
2.   Output: Recommended_model
3.    //Keep track of supply and demand
4.    If a broker is terminated, update the demand function
5.    If a resource is committed, update the supply function
6.
7.    If(Optimization-metric = metric-n)  // where n is one of the
                                          //available performance metrics
8.      {
9.       Domain-i: Assuming that there are i number of different
            domains dominated by different economic models for this
            particular optimization
10.         compute the gradient using the current Demand and Supply
            for Domain-i
11.         if the gradient sits in Domain-i, select the corresponding
            model of the domain
12.          Recommended_model = model corresponding to the Domain-i
13.         }
14.   If(current_model != Recommended_model)
15.         {
16.          Let the environment know about the Recommended_model
17.          let delay = 0;
18.          sendMessage(environment, delay, Tag.Recommended_model);
19.         }
```

Now, based on the optimization function as defined by the Grid provider, the agent manipulates the decision space. To do this, the agent uses our identification of domains of strengths. For example, if the optimization function is maximization of revenue and minimization of communication cost, the decision space is described as,

Decision Space = {BAR, CDA, CNP}

As these three economic models have been identified suitable for maximizing revenue considering communication cost in their respective domains compared to other models (Fig. 1.), these are considered as decision variables. However, at a specific time in the environment, the agent can choose only one decision variable. To facilitate the decision process, the agent then converts the decision space in terms of their respective domains. From Fig. 1. , the decision space can be rewritten as,

$$\text{Decision Space} = \{d \geqslant 1 \text{ and } ((s+46.06)/d) \geqslant 13.01,\ ((s+46.06)/d) < 13.01 \text{ and } ((s-1.93)/d) \geqslant 0.07,\ ((s-1.93)/d) < 0.07 \text{ and } s \geqslant 1\}$$

Where, s and d refer to the supply and demand respectively at any state in the environment. For a specific time, either one of the domains in the decision space will be true. As the agent is notified about every occurrence of supply or demand change in the environment, it can determine the true domain respective to s and d of any given state. Thereafter, the agent chooses the model corresponding to that domain as an input to the environment (*Pseudocode-3*). The nature of s and d in the decision space ensures the scalability of the domains identified; i.e. can be any value greater than or equal to 1. Therefore, if sometimes, there is any spike in supply/demand, the model would still be able to manipulate its decision space.

VI. Experiment

To evaluate our model, the simulation starts with a constant number of resources and continuously injects different number of users over times to obtain a dynamic variation of supply and demand. A new set of users is injected as soon as the previous set is finished. This scenario continues until the total number of resources finishes in the environment. For statistical need, all the simulations are carried out with five different samples and presented only their averages.

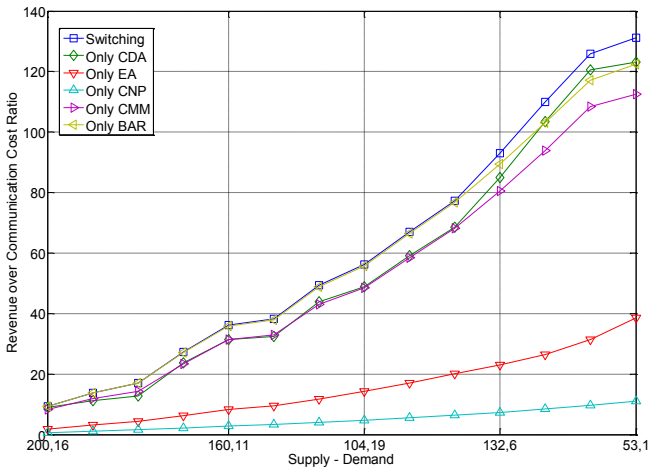


Fig. 3. Revenue over Communication Cost Comparison

Optimization Metric: *Revenue over Communication Cost*. The agent is informed beforehand about the optimization metric and then it dynamically decides which model to be used when and for what scenario.

Fig. 3. illustrates the results obtained during the optimization procedure for revenue over communication cost. The X-axis of the plot shows different supply-demand ratios at different times. The supply starts at 200 and keeps decreasing as occupied by the randomly injected users over time. For injecting users, the range (1, 20) is used to make the switching model in action (refer to Fig. 1.). The Y-axis represents the revenue over communication cost ratio for different models. Due to different units, the both parameters are normalized and then computed the ratios.

It can be observed that overall the switching model performs better compared to any other individual models. At each different scenario (along X-axis), the agent determines which model will be suitable to optimize the defined function. For example, when the environment notices that the current supply is 160 units and demand is 11 units, the agent computes the suitable protocol for this particular supply-demand ratio, which is BAR. Therefore, the switching model closely follows the trend of BAR here. This scenario continues until the supply approaches at 132 units. At this stage, the model determines CDA to be suitable. As the previous performance of BAR is now added to CDA, switching starts performing better even than the CDA itself. Because of switching between the models, the system is able to better utilize the relative strengths of the models and ultimately to optimize the function. Thus, the switching model outperforms all the five models individually. The optimization for other metrics such as time and social welfare has been omitted due to space constraint.

VII. Conclusions

The demand of Grid computing has been broadening day by day. Economic models are found compelling because they can cope with the dynamics of the Grid. This work has identified that different economic models are suitable for different scenarios where the scenarios are characterized in terms of supply and demand. Based on the identification about the strengths and weaknesses of different economic models, this paper proposed an optimization framework in a dynamic Grid environment. The paper described the framework in terms of dynamically switching from one model to another based on the models' domains of strengths. It further discussed the extended broker, resource and system's behavior in terms of an autonomous and learning environment. The paper described the agent and its role in the environment through automatically deciding which model to use at what scenario. Through an experimental analysis, it showed that our multi-model framework has successfully been able to optimize predefined objective functions better than any other individual models in the environment. Future work would evaluate effectiveness of the switching framework for Cloud computing platform.

References

[1] R. Buyya, D. Abramson, J. Giddy, and H. Stockinger, "Economic models for resource management and scheduling in Grid computing," *Concurrency and Computation: Practice and Experience,* vol. 14, pp. 1507-1542, 2002.

[2] R. Beck, M. Schwind, and O. Hinz, "Grid Economics in Departmentalized Enterprises," *J Grid Computing,* vol. 6, pp. 277-290, 2008.

[3] R. Buyya, D. Abramson, and S. Venugopal, "The Grid Economy," *Proceedings of the IEEE,* vol. 93, pp. 698-714, 2005.

[4] A. Haque, S. M. Alhashmi, and R. Parthiban, "A survey of economic models in grid computing," *Future Generation Computer Systems,* vol. 27, pp. 1056-1069, 2011.

[5] A. Haque, S. Alhashmi, and R. Parthiban, "An Inspiration for Solving Grid Resource Management Problems Using Multiple Economic Models," in *Economics of Grids, Clouds, Systems, and Services*. vol. 7150, Eds., K. Vanmechelen, J. Altmann, and O. Rana, Springer Berlin / Heidelberg, 2012, pp. 1-16.

[6] R. Wolski, J. S. Plank, J. Brevik, and T. Bryan, "G-commerce: market formulations controlling resource allocation on the computational grid," in *Proceedings 15th International Parallel and Distributed Processing Symposium*, 2001, p. 8 pp.

[7] Z. Tan and J. R. Gurd, "Market-based grid resource allocation using a stable continuous double auction," in *8th IEEE/ACM International Conference on Grid Computing*, 2007, pp. 283-290.

[8] M. Caramia and S. Giordani, "Resource allocation in grid computing: an economic model," *J. WSEAS Trans. Comp.,* vol. 3, pp. 19-27, 2008.

[9] M. Assuncao, W. Streitberger, T. Eymann, and R. Buyya, "Enabling the Simulation of Service-Oriented Computing and Provisioning Policies for Autonomic Utility Grids," in *Proceedings of the 4th International Workshop on Grid Economics and Business (GECON'07)*, Rennes, France, 2007.

[10] G. Stuer, K. Vanmechelen, and J. Broeckhove, "A commodity market algorithm for pricing substitutable Grid resources," *Future Generation Computer Systems,* vol. 23, pp. 688-701, 2007.

[11] D. C. Parkes and L. H. Ungar, "Iterative combinatorial auctions: theory and practice," 2000.

[12] L. Xing and Z. Lan, "A Method Based on Iterative Combinatorial Auction Mechanism for Resource Allocation in Grid Multi-agent Systems," in *International Conference on Intelligent Human-Machine Systems and Cybernetics, 2009. IHMSC '09*, 2009, pp. 36-39.

$iDMI$: A Novel Technique for Missing Value Imputation using a Decision Tree and Expectation-Maximization Algorithm

Md. Geaur Rahman and Md Zahidul Islam
Center for Research in Complex Systems (CRiCS), School of Computing and Mathematics
Charles Sturt University, Bathurst, NSW 2795, Australia.
Emails: {grahman, zislam}@csu.edu.au

***Abstract*—In this paper we present a novel technique called $iDMI$ that imputes missing values of a data set by combining a decision tree algorithm (DT) and an expectation-maximization (EMI) algorithm. We first divide a data set into horizontal segments through applying a DT algorithm such as C4.5, and then apply an EMI algorithm on each segment in order to impute the missing values belong to the segment. If all numerical attribute values of a record are missing then we impute them by the mean values of the attributes of the records belong to a segment where the record falls in, and thereby reduce the computational time complexity of $iDMI$ compare to an existing technique called DMI which calculate the mean value of an attribute by using all records of a data set. We evaluate the performance of $iDMI$ over three high quality existing techniques on two real data sets in terms of four evaluation criteria. Our initial experimental results, including several statistical significance analysis, indicate the superiority of $iDMI$ over the existing techniques.**

Keywords: Data pre-processing; data cleansing; missing value imputation; EM algorithm; Decision Trees

I. Introduction

Data mining techniques play a vital role for decision making process in an organization by extracting the hidden and useful patterns from large data sets. However, the data sets often suffer from erroneous (i.e. inaccurate/missing) data due to a variety of reasons including equipment malfunctioning, faulty data collection, transmission and propagation, and human errors and misunderstanding [1]. The data sets can contain around 5% or more erroneous data if an organization does not take extreme care during the data collection [2], [3]. The existense of erroneous data in data sets can adversely affect the accuracy of data mining techniques. Moreover, the use of data mining techniques such as clustering or classfication on the data sets having erroneous data can produce misleading results that may make the data sets less useful to the data miners [4].

Since the repetition of the data collection process is expensive and often impractical, it is essential to have an effective data cleansing framework in order to deal with the erroneous data. Imputation of missing values is one of the important tasks of data cleansing. For imputing missing values in the data sets a number of techniques have been proposed [1], [5]–[9]. However, the imputation accuracy of a technique relies on both the amount and the type/pattern of missing values [10], [11].

In this paper we adopt the following notations for all techniques. We consider a data set D_F as a two dimensional table with N records (or rows) and M attributes (or columns). The attributes of D_F can be of two types namely numerical (e.g. "1", "1.5") and categorical (e.g. "Bangladesh", "Australia"). Let $R = \{R_1, R_2, \ldots, R_N\}$ be the records of D_F and $A = \{A_1, A_2, \ldots, A_M\}$ be the attributes of D_F. Also let r_{ij} be the value of the ith record and the jth attribute. By $r_{ij} =?$ we mean the value of r_{ij} is missing. Again let $Q = \{R_1, R_2, \ldots, R_n\} \subset D_F$ be the set of records where all numerical attribute values are missing.

In order to impute the numerical missing value/s of a record R_i, an advanced approach called Expectation-Maximization algorithm (EMI) [7], [10] uses the mean values of all numerical attributes, and the correlations between attributes having missing values and attributes having available values in R_i. However, the technique does not work for a record $R_c \in Q$, where all numerical attribute values are missing, due to the unavailability of correlations between attributes having missing values and attributes having available values of R_c.

Since the correlations of attributes within a horizontal segment (that can be obtained by a leaf of a decision tree) of D_F are higher than the correlations of attributes of a whole data set D_F, a recent technique called DMI [1] first finds a set of leaves from D_F by using an existing decision tree algorithm such as $C4.5$ [12]. DMI then applies an EMI algorithm within each leaf for imputing numerical missing value/s. DMI imputes a missing value r_{ij} of a record $R_i \in Q$, where all numerical attribute values of R_i are missing, by considering the mean value of the attribute A_j for all available values of A_j from the whole data set D_F. It is shown that DMI performs better than EMI in terms of imputation accuracy and the improvement of data quality [1], [13].

Like other existing techniques, the imputation accuracy of DMI declines for a data set with the complex missing pattern (see Figure 2). Since almost 80% of attribute values are missing in a complex pattern [1], a record with a complex pattern can even lose all numerical attribute values. Therefore, a data set with a complex pattern contains a huge number of

records in Q where all numerical attribute values are missing (see Table I). One possible reason of dropping the imputation accuracy of DMI, for the complex pattern, could be the use of the whole data set for imputing the missing values of a record $R_c \in Q$. Therefore, there is a room in DMI for further improvement.

It is evident that the average similarity of the records belong to a leaf is higher than the average similarity of the records of a whole data set [9]. Therefore, we argue that the imputation accuracy of DMI can significantly be improved if it imputes the records of Q by using the attribute mean value of the records that belong to leaf, instead of using the whole data set D_F. Hence, in order to get better imputation accuracy, we propose a technique called $iDMI$ which imputes a record $R_c \in Q$ by considering the mean value of the records within a leaf where the record R_c falls in.

In this paper we have the following novel contributions: 1) For a record $R_c \in Q$, where all numerical attribute values are missing, we consider the mean value of the records of a leaf where R_c belongs to, 2) We carry out complete experiments that indicate the higher imputation accuracy of our proposed technique than the existing techniques, and 3) The execution time complexity of $iDMI$ is lower than DMI since DMI calculates the mean value by considering a whole data set whereas $iDMI$ considers only a leaf of the data set.

We evaluate the performance of our proposed technique $iDMI$ over the the performances of three high quality existing techniques namely DMI [1], EMI [7] and IBLLS [6] in terms of four evaluation criteria namely co-efficient of determination (R^2) [10], Index of agreement (d_2) [14], root mean squared error ($RMSE$) [10] and mean absolute error (MAE) [10] on two real data sets namely Chess and Yeast that are publicly available in UCI machine learning repository [15]. For simulating missing values we use four missing patterns namely Simple, Medium, Complex and Blended, four missing ratios: 1%, 3%, 5% and 10%, and two missing models namely Overall and Uniformly Distributed (UD). A detailed description about the missing patterns, missing models, and missing ratios is available in the literature [1], [3], [10]. The initial experimental results, including 95% confidence interval analysis and statistical t-test analysis, indicate that our proposed technique $iDMI$ performs significantly better than the existing techniques.

The organization of the paper is as follows. Some related works are presented in Section II. Section III presents our proposed technique. Experimental results are presented in Section IV and the conclusion of this paper is presented in Section V.

II. Background Study

A number of techniques have been proposed for imputing missing values [6]–[9], [16], [17]. An early approach deals with missing values by deleting the records having missing value/s [18]. However, the deletion of the records, having missing value/s, from a small sized data set generally reduces the usability of the data set for various statistical analysis. In addition, the statistical analysis results obtained from a data set having insufficient number of records, due to deletion, can produce misleading information [19].

In order to impute a missing value $r_{ij} \in R_i$ of an attribute A_j, another early technique uses the mean of all available values of the jth attribute in D_F [20]. However, in terms of data mining and statistical analysis, the mean imputation technique can often produce more misleading results than the simple record deletion approach [19].

For imputing a missing value r_{ij}, an advanced technique called k-Nearest Neighbour Imputation (kNNI) [21] first finds the (user-defined) k-most similar records of R_i from D_F. If A_j is categorical, the technique imputes r_{ij} by using the most frequent value of A_j within the k-Nearest Neighbour (k-NN) records. Otherwise if A_j is numerical, the technique uses the mean value of A_j for the k-NN records. It is indicated that the imputation accuracy of kNNI is better than the imputation accuracy of normal mean/mode imputation technique that calculates mean/mode from a whole data set D_F, instead of the k-NN records of D_F [22]. However, the computational complexity of kNNI is higher for a large data set D_F since the technique finds the k-NN records of each record having missing value/s by searching the whole data set. Moreover, in real life data sets, it is challenging to find a suitable value for k [21].

The value of k is determined automatically in a recent technique called IBLLS [6] which first creates an artificial missing value in R_i. For all possible values of k in D_F, it then imputes the missing value by using a Local Least Square (LLS) framework [6]. Since the actual value of the artificially created missing value is known, it calculates normalized RMSE for all k considering the actual value and the imputed value. The value of k which gives the best RMSE is chosen for imputing the real missing values of R_i. While imputing r_{ij}, it further divides the k-NN records vertically by considering the attributes having high correlations with A_j. It completely ignores the attributes having low correlations with A_j resulting a lower imputation accuracy of IBLLS for a data set having low number of numerical attributes [5], [9], [17].

III. Our Novel Imputation Technique - $iDMI$

We propose a novel technique called $iDMI$ that combines a Decision Tree (DT) algorithm (such as C4.5 [12]) and an Expectation Maximization (EMI) [7] algorithm for imputing missing values of a data set, D_F. Before we discuss the technique in details we first introduce the basic concepts.

For imputing numerical missing values, an existing technique DMI [1] uses an EMI algorithm on the horizontal segments (also known as leaves) of a data set D_F. It is expected to have higher imputation accuracy by DMI than EMI, since the correlations of attributes for the records that belong to a leaf are higher than the correlations of attributes for the records within the whole data set [1]. Note that EMI relies on the correlations between attributes, A_m, having missing values and the attributes, A_a, having available values for imputing numerical missing values. EMI does not impute the records of Q, where all numerical attribute values are missing,

due to unavailability of the correlations between A_m and A_a. DMI imputes the records in Q by using attribute mean of the whole data set D_F.

In real life data sets, it is common to have records Q, where all numerical attribute values are missing. For example, Fig. 1 shows the percentage of records with all numerical values that are missing in the CMC and Chess data sets, where we artificially simulate missing values by using four missing ratios, four missing patterns and two missing models [1]. Fig. 1 shows that the size of Q increases for a high missing ratio. We also present numerical analysis of Q in Table I which shows the number of records in Q for the CMC and Chess data sets based on a combination of 10% missing ratio, "Overall" missing model and "Complex" missing pattern.

In Fig. 1 and Table I, we see that both data sets contain a large number of records where all numerical attribute values are missing. For the CMC data set, the number of records having one or more missing values ($|D_I|$) is 243 (Column C of Table I). Out of 243, there are 83 records in Q where all numerical attribute values are missing as shown in Column E of Table I. The ratio of Q and D_I is alarming on both data sets. For example, the ratio for the CMC data set is 34.16% (see Column G of Table I).

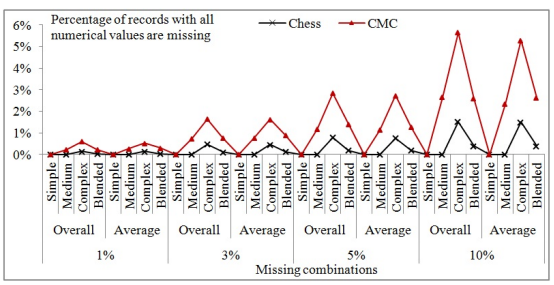
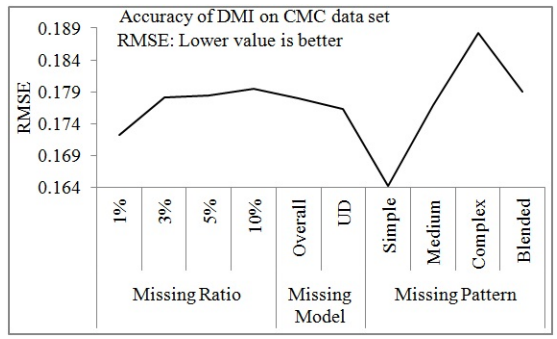


Fig. 1. Percentage of records with all numerical values are missing on CMC and Chess data sets

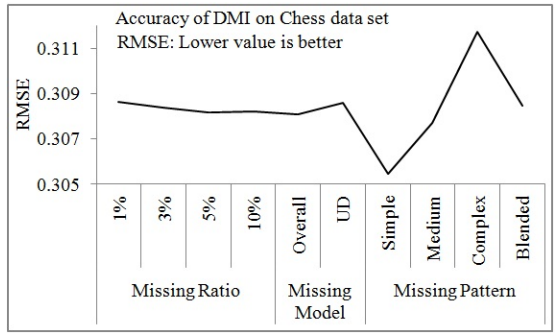
TABLE I
NUMBER OF RECORDS WITH ALL NUMERICAL VALUES ARE MISSING FOR CMC AND CHESS DATA SETS IN TERMS OF 10% MISSING RATIO, "OVERALL" MISSING MODEL AND "COMPLEX" MISSING PATTERN.

Data set D_F	Total Records or $\|D_F\|$	Records belong to D_I $\|D_I\|$	In % w.r.t. $\|D_F\|$	Records belong to Q $\|Q\|$	In % w.r.t. $\|D_I\|$	In % w.r.t. $\|Q\|$
A	B	C	$D = \frac{C}{B}$	E	$F = \frac{E}{B}$	$G = \frac{E}{C}$
CMC	1473	243	16.50%	83	5.63%	**34.16%**
Chess	28056	5608	19.99%	423	1.51%	7.54%

Since DMI imputes the records in Q with the mean values of the records of a whole data set D_F, the imputation accuracy of DMI can be lower due to the existence of a huge number of records in Q. We empirically analyze the concept on the CMC and Chess data sets as shown in Fig. 2. The figure demonstrates that DMI fails to perform well for complex pattern in both data sets. The reason of dropping the imputation accuracy of DMI for the complex pattern is imputation of the missing values of a record $R_c \in Q$, by considering the whole data set. Therefore, DMI has a room for further improvement.



(a) CMC data set



(b) Chess data set

Fig. 2. Accuracy of DMI in terms of $RMSE$ on CMC and Chess data sets.

We argue that, while imputing a record $R_c \in Q$ with all numerical attribute values being missing, the performance of DMI can significantly be improved if it considers the records that are belonging to a leaf (instead of the whole data set) since the average similarity of the records within a leaf is higher than the average similarity of the records of a whole data set [9]. Moreover, it is evident that the correlations of attributes for the records within a leaf are generally higher than the correlations of attributes for the records of a whole data set [1]. Therefore, for achieving a better imputation accuracy, we propose a technique $iDMI$ which imputes missing values of the records in Q by using the mean values of the attributes of the records belonging to a leaf instead of the whole data set. We now discuss our technique in details as follows.

We first divide a data set D_F into two sub data sets namely D_A and D_I where D_A contains records without missing values and D_I contains records with missing values. For each attribute $A_j \in A$ having missing values, we build a decision tree (DT) T_j through applying a decision tree algorithm such as $C4.5$ [12] on D_A by considering the attribute A_j as the class attribute. If A_j is numerical then we first generalize values of A_j on D_A into N_A categories where N_A is the square root of the difference between the maximum and minimum values of A_j. Note that for each T_j we have a set of logic rules $S_j = \{S_{j1}, S_{j2}, \ldots, S_{j|S_j|}\}$. For each logic rule S_{jl} of T_j, we generate a leaf L_{jl} by inserting records, from D_F, that satisfy the logic rule S_{jl}.

For a missing value $r_{ij} \in R_i$, of a record R_i in D_I, we first identify the leaf L_{jl}, using the T_j, where the record R_i

belongs to. If R_i falls in multiple leaves then we use any of them. If A_j is categorical then we consider the majority class value for the logic rule S_{jl} of DT T_j as the imputed value for r_{ij}. On the other hand, if A_j is numerical and R_i does not belong to Q then we impute r_{ij} by applying the EMI algorithm [7] on the leaf L_{jl}. Whereas, if A_j is numerical but R_i belongs to Q then we impute r_{ij} by the mean value of the jth attribute for the records that belong to the leaf L_{jl}.

We now explain the imputation procedure of $iDMI$ with an example as follows. We consider a toy data set D_F with 15 records and 4 attributes as shown in Table IIa. In the table all values are available initially. We then artificially create some missing values in the data set. Table IIb shows the data set D_F with missing values. We now divide D_F into D_I and D_A where D_I contains records with missing values and D_A contains records with all values are available. Since all numerical attribute values are missing for records "R1" and "R2", the records also fall in Q. Using D_A, we build a decision tree (DT) through applying the decision tree algorithm $C4.5$ [12]. The DT is shown in Fig. 3 where a decision node is represented by a circle and a leaf is represented by a rectangle. In the DT, Leaf 1 contains 5 records (R4, R6, R10, R12, R14) having the same value "MS" for the attribute "Edu." whereas Leaf 2 contains 8 records (R3, R5, R7, R8, R9, R11, R13, R15) having same value "PhD" in the attribute "Edu.".

Based on the decision node, we now assign the records of Q into the two leaves. Since $r_{R1,Edu.} = MS$, we assign "R1" into Leaf 1, and similarly "R2" is assigned in Leaf 2 (not shown in the figure). Note that for the 5 records of Leaf 1, the mean values of the attributes "Age" and "Salary" 25.8 and 85.2, respectively. We then impute $r_{R1,Age} = 25.8$ and $r_{R1,Salary} = 85.2$ as reported as bold font in Table IId. Similarly we impute $r_{R2,Age} = 43.0$ and $r_{R2,Salary} = 144.0$ by taking the mean values of the attributes "Age" and "Salary" from Leaf 2. Besides, we present the DMI imputed values of the missing values of "R1" and "R2" as italic font in Table IIc. Note that DMI imputes them by the mean value of the whole data set D_F (Table IIb).

Based on the illustration from the toy data set, it is easily understandable that the values imputed by iDMI (Table IId) are more closer to the actual values, shown in Table IIa, than the values imputed by DMI (Table IIc). This is also reflected in the experiments that are carried out on the real life data sets and presented in the next section.

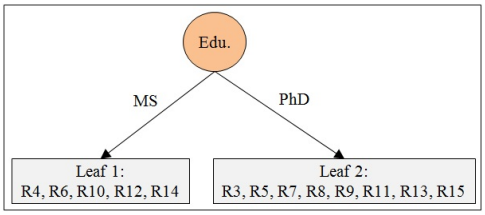


Fig. 3. A decision tree built from D_A.

TABLE II
A TOY DATA SET D_F

Rec.	Age	Edu.	Salary	Pos.
R1	27	MS	85	L
R2	45	PhD	145	P
R3	42	PhD	145	P
R4	25	MS	85	L
R5	50	PhD	146	P
R6	28	MS	85	L
R7	38	PhD	140	P
R8	43	PhD	147	P
R9	44	PhD	146	P
R10	25	MS	86	L
R11	42	PhD	142	P
R12	26	MS	84	L
R13	42	PhD	143	P
R14	25	MS	86	L
R15	43	PhD	143	P

(a) A toy data set D_F (Original)

Rec.	Age	Edu.	Salary	Pos.
R1	?	MS	?	?
R2	?	PhD	?	P
R3	42	PhD	145	P
R4	25	MS	85	L
R5	50	PhD	146	P
R6	28	MS	85	L
R7	38	PhD	140	P
R8	43	PhD	147	P
R9	44	PhD	146	P
R10	25	MS	86	L
R11	42	PhD	142	P
R12	26	MS	84	L
R13	42	PhD	143	P
R14	25	MS	86	L
R15	43	PhD	143	P

(b) D_F with missing values

Rec.	Age	Edu.	Salary	Pos.
R1	*36.4*	MS	*121.4*	*L*
R2	*36.4*	PhD	*121.4*	P

(c) Imputed by DMI

Rec.	Age	Edu.	Salary	Pos.
R1	**25.8**	MS	**85.2**	**L**
R2	**43.0**	PhD	**144.0**	P

(d) Imputed by LDMI

IV. Experimental Results and Discussion

We implement our novel technique iDMI and three existing techniques namely DMI [1], EMI [7] and IBLLS [6] in order to compare the accuracy of iDMI with the accuracies of the existing techniques on two real data sets namely Chess and CMC [15]. We compare the performances of the techniques on both large and small data sets. The Chess data has 28,056 records with 3 numerical and 4 categorical attributes. Besides, the CMC data set contains 1473 records with 2 numerical and 8 categorical attributes. Note that there are no record with natural missing values in the data sets. Therefore, we use all records of each data set in the experiment.

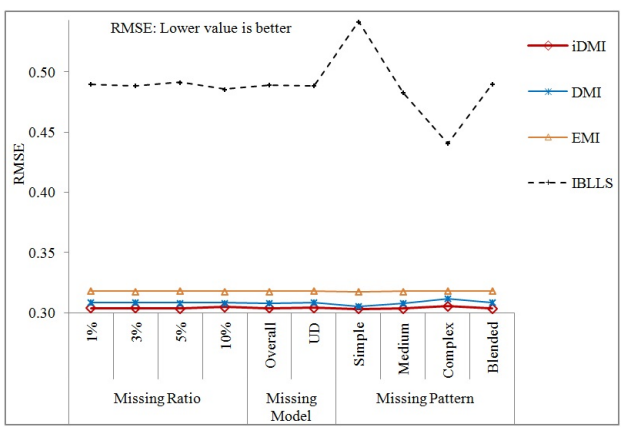
In the data sets, we then artificially create missing values which have been imputed by the four techniques. Since we know the actual values of the artificially created missing values, we can evaluate the imputation accuracy of the techniques by comparing the actual and imputed values. For simulating missing values, we use four missing patterns, four missing ratios and two missing models [1], [10]. So we have 32 missing combinations, and for each combination we create 10 data sets. Therefore, altogether we create 320 data sets for each natural data set.

Now we present the overall average performances of iDMI, DMI, EMI and IBLLS on CMC and Chess data sets in terms of R^2, d_2, $RMSE$ and MAE in Table III. Note that the outputs of both R^2 and d_2 can be between 0 and 1, where a higher value means a better imputation [9], [10]. On the other hand, the outputs of both $RMSE$ and MAE can be between 0 to ∞, where a lower value indicates a better imputation [9], [10]. In the table, a value is the average value of the accuracies of the 320 data sets. For example, the overall accuracy of iDMI on CMC data set for R^2 is 0.535 which is the average of the R^2 values of iDMI from the 320 data sets that are created from the CMC data set. Bold values indicate the best results among the techniques. It is clear from table that in both data sets iDMI performs better than other techniques in terms of all evaluation criteria.

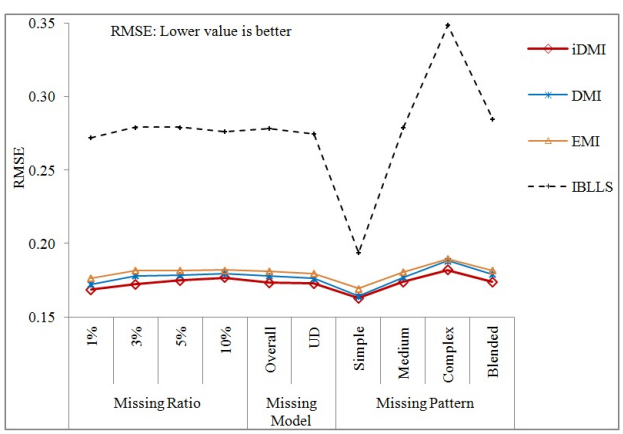
TABLE III
OVERALL AVERAGE PERFORMANCE ON CMC AND CHESS DATA SETS

Evaluation Criteria	CMC				Chess			
	iDMI	DMI	EMI	IBLLS	iDMI	DMI	EMI	IBLLS
R^2	**0.535**	0.514	0.498	0.269	**0.151**	0.150	0.097	0.065
d_2	**0.827**	0.811	0.799	0.699	**0.587**	0.495	0.432	0.429
$RMSE$	**0.173**	0.177	0.180	0.277	**0.304**	0.308	0.318	0.489
MAE	**0.133**	0.136	0.140	0.208	**0.258**	0.262	0.272	0.387

We also analysis the results in terms of missing patterns, missing ratios, and missing models as shown in Fig. 4. Due to space limitation, in the figure we present the results only for $RMSE$. The performance of iDMI based on $RMSE$ is better than the performances of DMI, EMI, and IBLLS on Chess data set (Fig. 4(a)) and CMC data set (Fig. 4(b)). It is notable that for the complex pattern iDMI improves the accuracy compare to DMI since iDMI imputes the records, with all numerical values are missing, by using the mean value of the records belong to a leaf instead of using the mean value of the records within a whole data set. We also get higher performance for iDMI over other techniques in terms of R^2, d_2, and MAE.



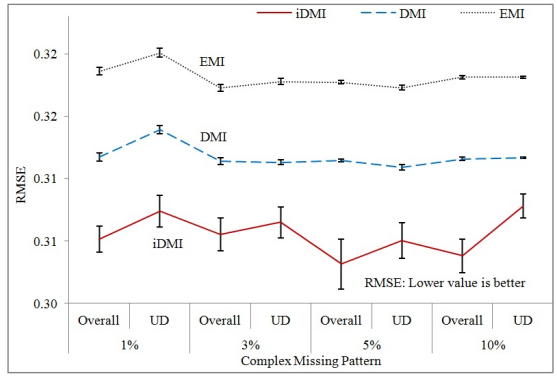
(a) Chess data set



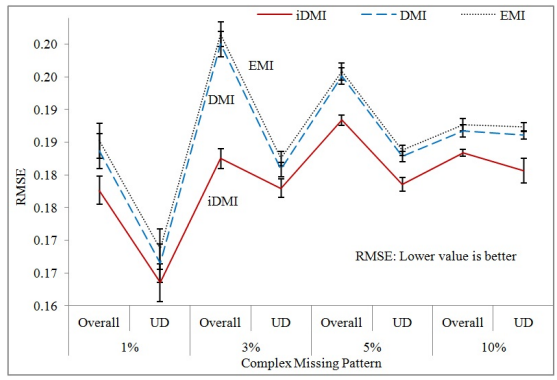
(b) CMC data set

Fig. 4. Aggregated performance on Chess and CMC data sets..

We use several statistical analysis including confidence interval and t-test in order to examine the goodness of our proposed technique. Since IBLLS performs the worst among the four techniques, for the complex pattern we present a 95% confidence interval analysis for iDMI, DMI, and EMI in terms of $RMSE$ in Fig. 5. It is clear from the figure that iDMI has higher average value than other technqiues for all 8 combinations in Chess data set (Fig. 5(a)) and in CMC data set (Fig. 5(b)).



(a) Chess data set



(b) CMC data set

Fig. 5. 95% confidence interval analysis in terms of $RMSE$ on Chess and CMC data sets.

In Fig. 6 we present a t-test analysis for all techniques based on all 32 missing combinations of all data sets. The figure shows that iDMI performs significantly better than other techniques at $p = 0.05$ based on all evaluation criteria for the Chess and CMC data sets except those marked by the arrows. The t-values are higher than the t(ref) values.

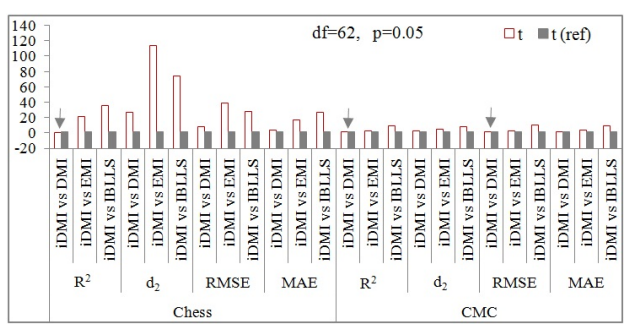


Fig. 6. t-test analysis on two data sets.

We now analyze the percentage of the combinations (out

of 64 missing combinations for the two data sets) where a technique performs the best among the four techniques. iDMI performs the best in 100% combinations in terms of $RMSE$, R^2, d_2 and MAE.

The experimental results demonstrate the effectiveness of iDMI since the technique improves imputation accuracy significantly by the use of the mean value of each numerical attribute of the records belong to a leaf in order to impute the records with all numerical values are missing. iDMI outperforms the others in terms of all evaluation criteria for both Chess and CMC data sets.

Table IV presents the average execution time (in milliseconds) for 320 data sets for the Chess and CMC data sets. The experiments are carried out by using two different machines where Machine 1 is configured with 4 $\times$ 8 core Intel E7-8837 Xeon processors and 256 GB RAM, whereas Machine 2 is configured with Intel Core i5 processor having speed 2.67 GHz and 4 GB RAM. However, the experiments on a data set are done by using the the same machine for all techniques. iDMI takes less time than IBLLS on both data sets, whereas it takes more time than EMI to achieve a significantly better quality imputation. Since iDMI calculates mean values of the attributes by considering the records belong to a leaf (which contains much lower number of records than a whole data set), it requires lower computational time than DMI in both data sets.

TABLE IV
AVERAGE EXECUTION TIME (IN MILLISECONDS) OF DIFFERENT METHODS ON THE TWO DATA SETS.

Data set	iDMI	DMI	EMI	IBLLS	Machine used
Chess	97,844	153,915	8,667	15,537,849	Machine 1
CMC	38,509	40,195	469	233,994	Machine 2

V. CONCLUSION

Data set often have records where all numerical attribute values are missing while one or more categorical attribute values are available. In this paper, we propose a novel technique called $iDMI$ which imputes those records by using the mean values of the attributes of the records belonging to the leaf where the record (with the missing values) falls in. The effectiveness of considering a leaf instead of the whole data set is reflected in our initial experimental results. The experiments are carried out on two real data sets in order to compare the performance of $iDMI$ with three high quality existing techniques namely DMI, EMI, and IBLLS in terms of four evaluation criteria namely R^2, d_2, $RMSE$, and MAE. $iDMI$ outperforms other techniques on the Chess and CMC data sets in terms of all evaluation criteria (Table III). According to the t-test in Fig. 6, $iDMI$ performs significantly better than other techniques at $p = 0.05$. Table IV shows that $iDMI$ requires less computational time than DMI and IBLLS since $iDMI$ uses a leaf, instead of the whole data set, for the calculation of the mean value. Our future plans include extensive experiment of $iDMI$ on more real data sets, and further improvement of $iDMI$ for achieving higher imputation accuracy.

REFERENCES

[1] M. G. Rahman and M. Z. Islam, "A decision tree-based missing value imputation technique for data pre-processing," in *Australasian Data Mining Conference (AusDM 11)*, ser. CRPIT, vol. 121. Ballarat, Australia: ACS, 2011, pp. 41–50. [Online]. Available: http://crpit.com/confpapers/CRPITV121Rahman.pdf

[2] J. Maletic and A. Marcus, "Data cleansing: Beyond integrity analysis," in *Proceedings of the Conference on Information Quality*. Citeseer, 2000, pp. 200–209.

[3] M. G. Rahman, M. Z. Islam, T. Bossomaier, and J. Gao, "Cairad: A co-appearance based analysis for incorrect records and attribute-values detection," in *Neural Networks (IJCNN), The 2012 International Joint Conference on*. Brisbane, Australia: IEEE, june 2012, pp. 1 –10.

[4] J. Han and M. Kamber, "Data mining: Concepts and techniques," *The Morgan Kaufmann Series in data management systems*, vol. 2, 2000.

[5] M. G. Rahman and M. Z. Islam, "A novel framework using two layers of missing value imputation," in *Australasian Data Mining Conference (AusDM 13)*, ser. CRPIT, vol. 146. in press, Canberra, Australia: ACS, 2013.

[6] K. Cheng, N. Law, and W. Siu, "Iterative bicluster-based least square framework for estimation of missing values in microarray gene expression data," *Pattern Recognition*, vol. 45, no. 4, pp. 1281–1289, 2012.

[7] T. Schneider, "Analysis of incomplete climate data: Estimation of mean values and covariance matrices and imputation of missing values," *Journal of Climate*, vol. 14, no. 5, pp. 853–871, 2001.

[8] X. Zhu, S. Zhang, Z. Jin, Z. Zhang, and Z. Xu, "Missing value estimation for mixed-attribute data sets," *Knowledge and Data Engineering, IEEE Transactions on*, vol. 23, no. 1, pp. 110–121, 2011.

[9] M. G. Rahman and M. Z. Islam, "Missing value imputation using decision trees and decision forests by splitting and merging records: Two novel techniques," *Knowledge-Based Systems*, vol. 53, pp. 51 – 65, 2013. [Online]. Available: http://dx.doi.org/10.1016/j.knosys.2013.08.023

[10] H. Junninen, H. Niska, K. Tuppurainen, J. Ruuskanen, and M. Kolehmainen, "Methods for imputation of missing values in air quality data sets," *Atmospheric Environment*, vol. 38, no. 18, pp. 2895–2907, 2004.

[11] D. Rubin, "Inference and missing data," *Biometrika*, vol. 63, no. 3, pp. 581–592, 1976.

[12] J. R. Quinlan, "Improved use of continuous attributes in C4.5," *Journal of Artificial Intelligence Research*, vol. 4, pp. 77–90, 1996.

[13] M. G. Rahman and M. Z. Islam, "Data quality improvement by imputation of missing values," in *International Conference on Computer Science and Information Technology (CSIT-2013)*, Yogyakarta, Indonesia, 2013, pp. 82 – 88.

[14] C. Willmott, "Some comments on the evaluation of model performance." *Bulletin of the American Meteorological Society*, vol. 63, pp. 1309–1369, 1982.

[15] A. Frank and A. Asuncion, "UCI machine learning repository [online available: http://archive.ics.uci.edu/ml]," 2010, accessed July 7, 2013. [Online]. Available: http://archive.ics.uci.edu/ml

[16] T. H. Bø, B. Dysvik, and I. Jonassen, "Lsimpute: accurate estimation of missing values in microarray data with least squares methods," *Nucleic acids research*, vol. 32, no. 3, pp. e34–e34, 2004.

[17] M. G. Rahman and M. Z. Islam, "kdmi: A novel method for missing values imputation using two levels of horizontal partitioning in a data set," in *The 9th International Conference on Advanced Data Mining and Applications (ADMA 13)*. in press, Hangzhou, China: Springer, 2013.

[18] A. Derjani Bayeh and M. J. Smith, "Effect of physical ergonomics on vdt workers' health: a longitudinal intervention field study in a service organization," *International Journal of Human-Computer Interaction*, vol. 11, no. 2, pp. 109–135, 1999.

[19] J. Osborne and A. Overbay, "Best practices in data cleaning," *Best Practices in Quantitative Methods*, pp. 205–213, 2008.

[20] J. L. Schafer and J. W. Graham, "Missing data: our view of the state of the art." *Psychological methods*, vol. 7, no. 2, p. 147, 2002.

[21] G. Batista and M. Monard, "An analysis of four missing data treatment methods for supervised learning," *Applied Artificial Intelligence*, vol. 17, no. 5-6, pp. 519–533, 2003.

[22] C. Liu, D. Dai, and H. Yan, "The theoretic framework of local weighted approximation for microarray missing value estimation," *Pattern Recognition*, vol. 43, no. 8, pp. 2993–3002, 2010.

Handwritten Bangla Digit Recognition Employing Hybrid Neural Network Approach

Tahsina Hashem*, Mohammad Asif† and Md. Al-Amin Bhuiyan‡
*†Department of Computer Science and Engineering, Jahangirnagar University, Savar, Dhaka-1342
*Email: tahsinahashem@gmail.com
†Email: asif_pogo@yahoo.com
‡Department of Computer Engineering, King Faisal University, Alahssa 31982, Saudi Arabia
‡Email: mbhuiyan@kfu.edu.sa

***Abstract*—Handwritten Bangla digit recognition is one of the most attractive area for researchers who have interest in image processing and pattern recognition field. In our everyday activities like bank check identification, passport and document analysis, number plate identification and especially in our postal automation service, recognition of handwritten digits plays a significant role. That's why a rich body of literature already has been published in this area. But most traditional techniques are generally based on complex feature extraction approach that introduces a great overhead in recognition tasks. Recently a novel approach Back-propagation algorithm is used for recognition which simplifies the recognition process but the main drawback is, the network takes lots of iterations to converge. This paper addresses a faster and efficient Hybrid Neural Network Solution called (BAM+BPNN) which is a combination of Bidirectional Associative Memory and Back-Propagation Neural Network. BAM is used for dimensional reduction and BPNN is used to train the neural network with the set of input patterns for acquiring separate knowledge of each digit. This research can take a decision that Hybrid Neural Network algorithm (BPNN with BAM) takes less iteration to train and less time to recognize digits than Back-propagation algorithm (BPNN). Experimental study shows the effectiveness of our proposed technique.**

Keywords*—*Hybrid Neural Network, Back-propagation, BAM, Handwritten Bangla Digit

I. Introduction

Digit Recognition refers to the process of translating the images of handwritten, typewritten or printed digits into a standard format. So that everyone can understand and use it for different purposes such as editing, searching/indexing and especially in postal automation service. Every literate human has his own manner of writing styles. Many peculiarities may also exist in different writing styles. Thus the recognition process becomes very much complicated by noisy inputs, distorted images and differences between type phases, sizes and fonts.

Bengali is one of the most spoken languages (ranked seventh [1]) in the world and is the native-tongue of 99% citizens of Bangladesh. A number of researches already has been conducted with feature extraction [2]–[7] and neural network [8]–[10] based handwritten Bangla digit recognition problem. Complex feature extraction approaches have showed good performance in digit recognition. But in this process, all the features for each digit are need to be collected, which introduces a great overhead in recognition tasks. Morshed et al. [9] and some other researchers [4], [5], [10] used the novel approach Back-propagation Neural Network algorithm [11], which successfully overcomes the limitations of feature extraction approach, but the major drawback is it requires a large number of iterations to converge.

Recently, Sanjit et al. presented a hybrid neural network solution [12] for Bangla character recognition which combines local image sampling and artificial neural network. We have utilized their idea and modified their algorithm to make our hybrid neural network solution (BAM + BPNN) faster and efficient.

In summary the contribution of this research are as follows:

- A Hybrid Neural Network Solution (***B**iderectional **A**ssociative* **M***emory and* ***B**ack-**P**ropagation **N**eural Network*) called **(BAM+BPNN)** has been proposed to recognize handwritten Bangla digits. Employing this algorithm we have achieved our main motive of reducing number of iterations as well as making the system robust and efficient.
- The processed image of Bangla digits have been directly taken rather than using problem dependent feature extraction process.
- To avoid loss of information of each digit characteristics, we do not input directly the smaller image to BPNN. Rather than we have used BAM to reduce dimension of the digit image. Then the output of BAM is used as input to BPNN.
- The proposed scheme has effectively used Multiple Training Encoding Strategy to overcome the initial limitations of BAM and later strengthen the training phase.
- A number of experiments have validated the performance of our algorithm.

The remainder of this paper is organized as follows:

Section II presents the proposed algorithm which is explained elaborately with adequate examples. Section III focuses on the performances of the algorithm with respect to a variety of parameters. Section IV draws the conclusions of the paper.

II. Overview of Proposed System

In this section we give a brief overview of our robust Hybrid Neural Network Approach to recognize Bangla digits. Our Hybrid Neural Network Approach can be applicable to any field for the purpose of recognition of handwritten Bangla digits. Here in this paper, we highlight it's application to postal automation service. Each of the phase is designed and implemented very carefully to make our recognition system more faster and efficient than previous research. The overall process of digit recognition is done into two major phases:

1) Image Processing
2) Hybrid Neural Network Approach

In the following subsections we will describe the phases listed above in a more details.

A. Image Processing

This is the first phase of our system. The goal of image processing is to extract the Bangla digits from the envelope and then convert these extracted digits into a binary format for recognition. This phase is very much essential for removing noise and for minimizing the differences between type phases, sizes and fonts of different handwriting styles. In our research all the digit samples are collected from Bangladesh mail envelope. Fig.1 shows some sample of handwritten digits.

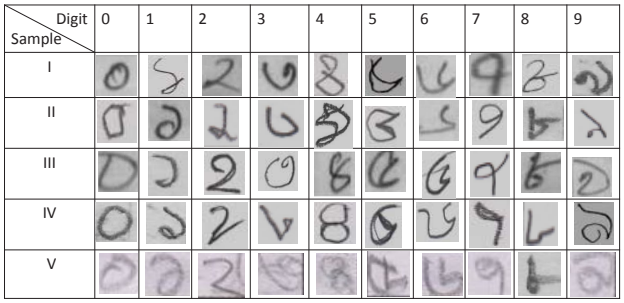


Fig. 1: Sample of Bangla Handwritten Digits

Image processing operations are performed through following steps:

1) **Scanning Image:** At first Bangladesh postal envelopes are scanned. Four digit postal codes are written in the box at lower right corner of the envelope. The digits were written by different people. So these digits vary in shape and size.
2) **Block Detection and Clipping:** The frame or block containing 4 digits postal codes is 15 pixels wide and 118 pixels long. Heuristic approach is used to clip the desired part. Clipping is done from fixed pixel position. The position is fixed by several trial and error methods. In this research, the position from which lower right portion is clipped: X=453, Y=256. So, 80% pixels are removed from envelope images. As a result, the probability of occurring error is reduced.
3) **Segmentation:** After detection of postcode frame, horizontal and vertical segmentation is done to finally extract each digit from the envelope.

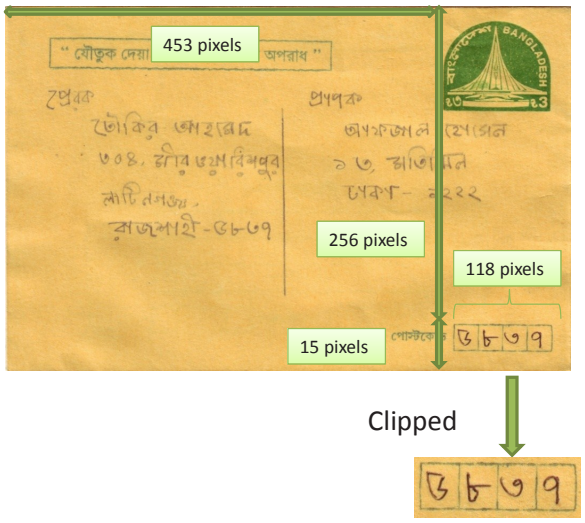


Fig. 2: Block Detection and Clipping Image

4) **Noise Reduction:** The image may contain noisy pixels. Such noises are removed for good performances even though neural networks has noise tolerance capability.
5) **Normalization and Scaling:** The segmented digits are then normalized to 16 by 16 pixels. Aspect Ratio Adaptive Normalization (ARAN) technique is used for normalization.
6) **Binarizing and Thinning:** Using Binarization the images are converted into matrices containing 0s for insignificant pixels and 1s for significant pixels. This research has applied Otsu's Method [13] for selecting the thresold value of the image. Based on this thresold value the image is converted to binary image. It is noticed that the thickness of the stroke is varying because different people use different writing tools. Since the thickness of the stroke influences the recognition result, thinning algorithm has been employed [14] to process the binary images of digits and get the same thickness of the stroke. So that all the thin images will have width 1 pixel.
At last the binary image is converted into binary digit matrix which consists of 0s and 1s. If the pixel is black, it becomes to 1 otherwise it is 0.

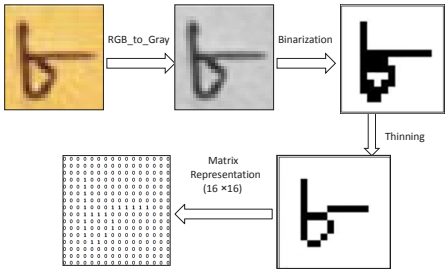


Fig. 3: Binarizing and Thinning

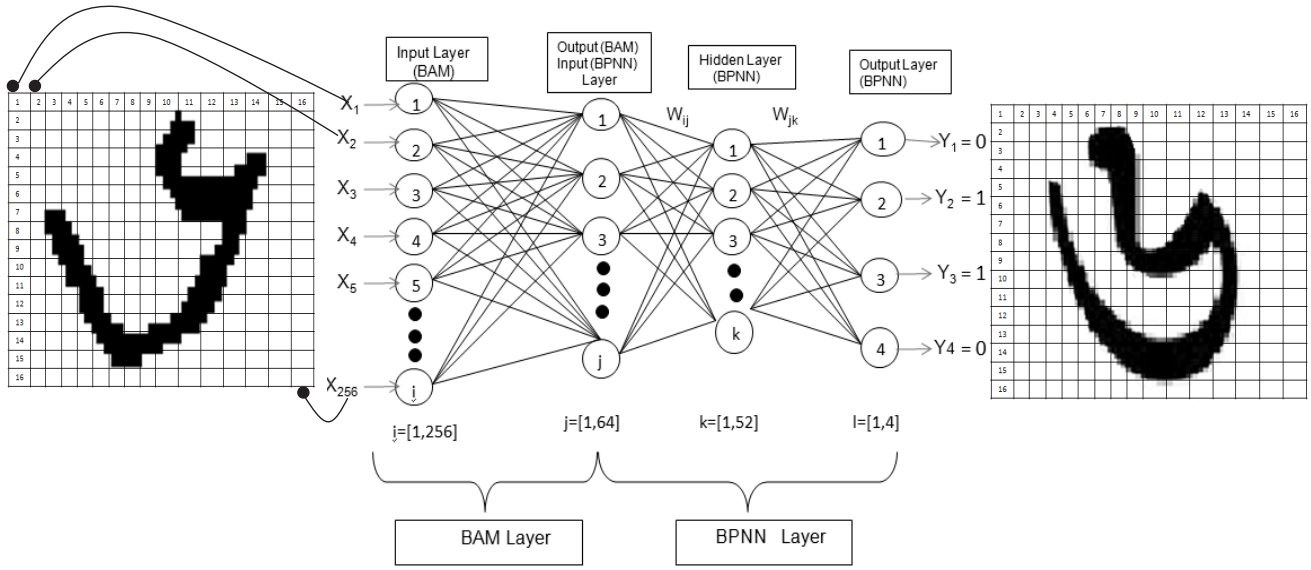


Fig. 4: Hybrid Neural Network Topology

B. Hybrid Neural Network Approach

The second stage of digit recognition uses the proposed hybrid neural network approach. A hybrid neural network is composed of two neural networks. One is **B**ack-**P**ropagation **N**eural **N**etwork (***BPNN***) and the other is **B**idirectional **A**ssociative **M**emory (***BAM***). The network architecture for the hybrid neural network is shown in Fig.4. The recognition process is divided into two phases:

1) Training Phase
   - Dimensional Reduction using BAM
   - Training Using BPNN
2) Recognition Phase

Each of the phase is discussed in details in the following subsections.

1) Dimensional Reduction Using BAM: Before training with Back-propagation neural network the dimension of the digit images are reduced. Employing **KOSKO's** hetero-associative correlator BAM, stored pattern pair are recalled even in presence of noise. The basic idea behind the BAM is to store pattern pairs for each digit. So that when M-dimensional vector **X** from set **A** is presented as input, BAM recalls N-dimensional vector **Y** from set **B**, where $M \geq N$.

But the limitation of BAM is it does not give guarantee to retrieve pattern pairs, when the stored pairs are at local minima. So in addition **WANG's** Multiple Training Encoding Strategy has been employed (which is an enhancement of the encoding strategy proposed by KOSKO) to make this research robust that ensures correct recall of al types of pattern pairs.

We have taken binary digit matrix as input matrix and employed BAM strategy to reduce dimension of the input digit matrix.

At first we have defined $m \times m$ digit matrix as $1 \times M$ digit input matrix of BAM which act as input vector **X** (where, $M = m \times m$ = Dimension of Input Digit Matrix). And the desired output is defined as $1 \times N$ digit matrix which is the representative of $n \times n$ digit matrix which act as output vector **Y** (where, $N = n \times n$ = Dimension of Output Digit Matrix). Then BAM generalized correlation matrix ($M \times N$) is constructed using WANG's Multiple Training Encoding Strategy. Finally this weight matrix is used to retrieve the stored pattern which is associated with the input pattern. As $m > n$, so dimension is reduced from $m \times m$ digit matrix to $n \times n$ digit matrix.

For example, consider the case m=16 and n=8. BAM recalls 8×8 Output Vector **Y** from 16×16 Input Vector **X** using 256×64 BAM correlation weight matrix. As a result, dimension is reduced from 16×16 digit matrix to 8×8 digit matrix. Fig.5 shows an example for sample digit THREE.

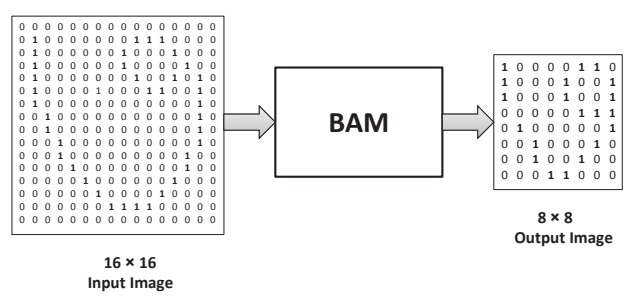


Fig. 5: Retrieving Associated Pattern Using BAM for Bangla Digit THREE

Algorithm 1 provides the pseudocode of the dimension reduction using BAM.

2) Training Using Back-Propagation Algorithm: The neural network is trained using back-propagation algorithm with the set of input patterns of the digits to learn the patterns corresponding to each Bangla digit.

Algorithm 1 Reduce Dimension Using BAM

1: Desired inputs ← binary matrix representation of each digit having larger dimension
2: Desired outputs ← binary matrix representation of each digit having smaller dimension
3: Calculate BAM weight matrix using Multiple Training Encoding Strategy for all associated pattern pairs
4: Save the BAM weight matrix for retrieving the associated pattern/digit pair

Algorithm 2 provides the pseudocode of the training process of hybrid neural network (***BAM+BPNN***). For each particular input pattern, algorithm 2 calls algorithm 3 to retrieve its associated output pattern from BAM. This associated pattern is defined as desired input of BPNN. The desired output of BPNN is a binary value of input decimal digit which consists of 4 neurons to classify 10 digits (0 to 9) in a binary format from 0000 to 1001. After initialization, training is continued until desired accuracy rate is achieved.

Algorithm 2 Training Hybrid Neural Network (***BAM+BPNN***)

1: **for** $each\ of\ the\ digit$ **do**
2: Input Pattern ← the binary matrix representation of digit having larger dimension.
3: Desired input of BPNN ← retrieve AssociatedPattern (Input Pattern)
4: Desired output of BPNN ← the binary form of decimal digit (4 neurons)
5: **while** $sumsq_error > 0.001$ **do**
6: Train the neural network using Back-propagation algorithm
7: **end while**
8: Save the finally trained weights and bias values for recognition
9: **end for**

Algorithm 3 retrieve AssociatedPattern (Input Pattern)

1: Output Pattern (smaller dimension) ← retrieve associated output of the Input Pattern (larger dimension) using saved BAM weight matrix
2: **return** Output Pattern

3) Recognition Phase: To finally recognize digit we input an unknown $m \times m$ digit matrix and retrieve it's associated stored pair $n \times n$ digit matrix using BAM (where $m > n$). This $n \times n$ digit matrix is defined as input of BPNN. During the recognition phase, no learning takes place i.e., weight matrices are not changed. There is no backward stage, only feed forward stage. Trained BPNN weight matrices are used to calculate the actual output in a binary format.

Algorithm 4 provides the pseudocode of the recognition process to identify the handwritten Bangla digits.

III. Experimental Results

This section presents the performance of the proposed system. In order to evaluate the effectiveness of the proposed system, various experiments were carried out for different

Algorithm 4 Recognize Handwritten Bangla Digit

1: Input Pattern ← binary matrix representation of an unknown pattern/digit having larger dimension
2: Desired input of BPNN (Output Pattern) ← retrieve AssociatedPattern (Input Pattern)
3: Saved trained weights and bias values are set to Back Propagation Neural Network to initialize
4: Actual output of the BPNN ← calculate using the back-propagation feed forward stage which is in a binary format.
5: Recognized digit ← Conversion (binary to decimal)

real captured images of Bangladesh postal envelope. The experimental results will be discussed in brief with respect to execution time and memory usage.

In order to justify the performance, 1400 digits were obtained from the real letters written by Bangladeshis. The samples set is a mixture of good and indistinct, obscure hand written digits to evaluate the proposed scheme effectively. Randomly 800 digits are selected as the training set and the rest of 600 digits are used as the test set.

In this research, we consider different combination of matrix size i.e. 10×10, 12×12, 16×16. If each digit is 16×16 pixels image, then the digit has a feature matrix of 256 elements in it. Each element is nothing but binary values (0 and 1). Hence, in the experiment the number of neurons in input layer of BAM is 256 and the number of neurons in output layer of BAM is 64 which in turn act as input layer for BPNN. The number of neurons in hidden layer is 52 (80% of input layer). Finally the number of neurons in output layer is 4 that generate 16 distinct binary values. However, in order to uniquely recognize the 10 decimal digits correctly we used the first 10 binary patterns only.

The training algorithm depends on various network parameters which are selected by observing the performance of the hybrid network. The values of network parameters are listed in table I.

TABLE I: Values of Network Parameters (16×16 Digit Sample)

Parameters	Values
No. of layers (BAM)	2 (1 Input & 1 Output Layer)
No. of input neurons in input layer (BAM)	16 × 16 = 256
No. of output neurons in output layer (BAM)	8 × 8 = 64
Activation function (BAM)	Y^{Sign}=1 (if X ≥ 37); 0 (if X < 37)
Momentum factor (BPNN)	0.95
Learning rate (BPNN)	0.2
Weight and Bias Initialization range (BPNN)	[-2.4/64,2.4/64]
No. of layers (BPNN)	3 (1 Input, 1 Hidden & 1 Output Layer)
No. of input neurons in input layer (BPNN)	8 × 8 = 64
No. of neurons in hidden layer (BPNN)	52 (80% of maximum no. of inputs)
No. of output neurons in output layer (BPNN)	4
Activation function (BPNN)	$Y^{Sigmoid}$=(1 / (1 + e^{-X}))
Error rate (Sum Squared Error)	0.001
Average number of epochs	1676

The sample output of the recognized digit 2 is shown in

Fig.6.

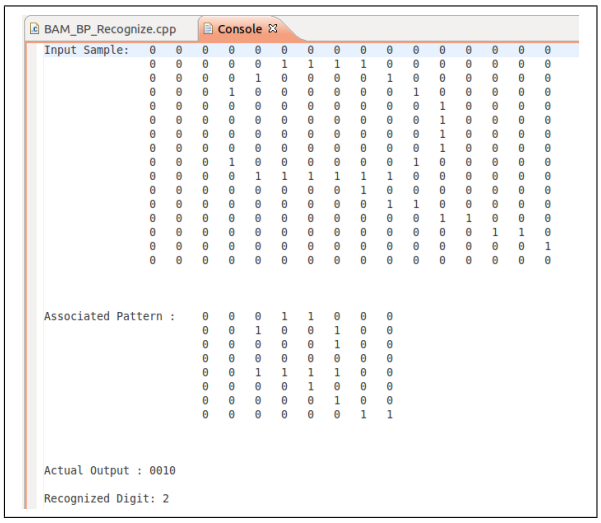


Fig. 6: Output of Test Image (Digit '2')

A. Error Minimization

Training is continued until the error comes to a minimum level. The error decreasing rate is graphically compared (%) between BPNN and BPNN with BAM with respect to iteration in the following Fig.7.

At first the task has been implemented using the BPNN algorithm and secondly, BAM and BPNN algorithm has been merged to train and recognize the Bangla Numeral. An error level is chosen (suppose 0.001 in percentage) to stop the training process. The error is calculated using sum of squared method. The number of iteration depends on the number of neurons in hidden layer and also on what algorithm has been used for the training process.

The performance of the network has been simulated by varying number of neurons in hidden layer from 50% to 90% with respect to input neurons in input layer. The training curve indicating the gradual reduction in error over several iterations due to our back-propagation (BPNN) learning algorithm. It is clear that the training curve converges more rapidly when 80% of input neurons (256) has been used in hidden layer. The number of iteration is 6360 when 50% of input neurons has been used in hidden layer but it decreases to 4070 when 80% of input neurons has been used in hidden layer. Finally when hybrid neural network (BPNN with BAM) approach has been used for the same percentage of hidden layer that is 80% of input neurons (64), the iteration decreases from 4070 to 1676. So, this research can take a decision that the considered hybrid neural network (BPNN with BAM) takes less iteration to train and less time to recognize digits than BPNN.

This research has been implemented in C++ on Linux platform with 3GB RAM. The performance of the algorithm has been tested using the following measures: recognition rate, error rate and rejection rate. Fig. 8 shows the bar diagram of the individual performance of different digits. From the figure it is seen that the proposed system achieves an average accuracy

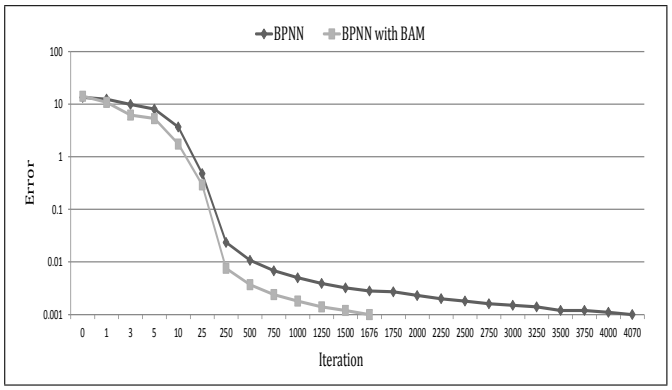


Fig. 7: BPNN vs BPNN with BAM (80% Hidden Layer)

rate (Recog.) of 91.1% and average error rate (Err.) and average rejection rate (Reject.) is 5.5% and 3.4% respectively. Highest accuracy rate is achieved for digit "eight" and "four". The mistaken recognition or rejection is due to variability of handwritings as well as bad writing.

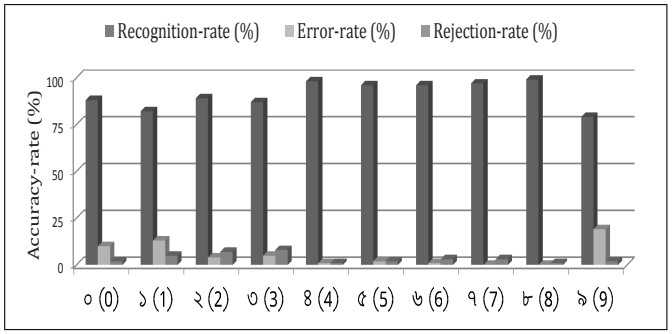


Fig. 8: Accuracy Rate of Each Digit

From the experiments it is noted that, our proposed scheme gives satisfactory outputs on recognition of indistinct handwritten digits. But sometimes it fails to recognize some special samples according to writers intension as because those samples are very much confusing. Fig.9 reflects some of such scenarios. The most confusing numeral pair is zero and three. Second confusing pair is one and two. Third confusing pair is one and nine. Therefore, error rate or rejection rate for these digits is higher. So the neural network needs to be trained using more samples especially for these digits.

We have also taken same noisy Bangla digits. Then we have observed the performance of our hybrid network on these noisy samples. Fig.10 shows the rate of error with different noise level. From the figure, it is noted that the recognition rate decreases with the increase of noise level.

Experimental results demonstrate that the presented system performs very well. It meets the response time as well as the accuracy requirements.

B. Cost Analysis

The performance of the hybrid network (BAM+BPNN) has been measured with respect to memory usage and execution

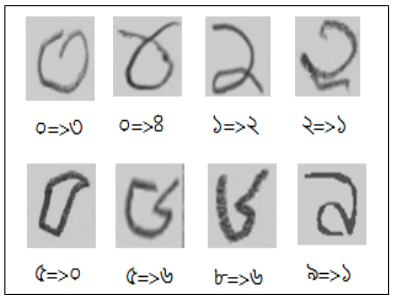


Fig. 9: Examples of some confused and indistinct handwritten Bangla digits

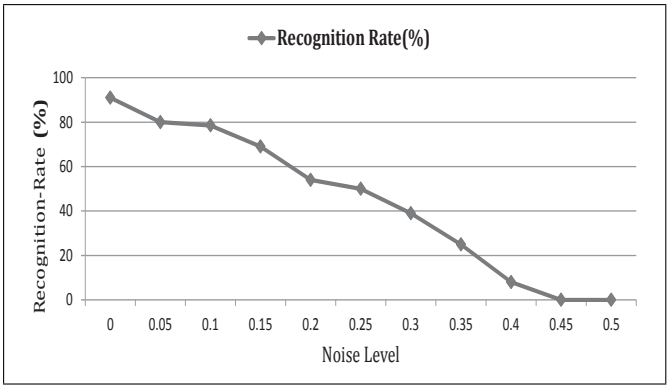


Fig. 10: Performance Evaluation on Noisy Digit Recognition

time. For the context of execution time, it has been observed that the training time is 21.14s and response time is 0.23s. Most of the time has been needed for training using back-propagation learning algorithm. But this training time is comparatively much shorter than previous research.

Now come to the memory consumption area. BAM is used in addition with BPNN for dimension reduction to make the recognition system faster. Although it requires some extra space to store the BAM correlation weight matrix for associating pattern pairs but it is the main reason that reduces the number of neurons in input layer of BPNN which in turn makes the training process faster.

IV. Conclusion

In this paper, we have proposed a hybrid neural network solution (BAM+BPNN) for the recognition of Bangla digits. Our main objective was to make the recognition system faster and efficient simultaneously. The back-propagation neural network continues to iterate until desired output is achieved. Hence we have merged the back-propagation and BAM algorithm. Moreover, to ensure correct recall of all types of pattern pairs multiple training encoding strategy has been employed in BAM. The performance is at satisfactory level since recognition time (0.23s) is minimized drastically and it meets the accuracy requirements (90%-95%) as well even on recognition of noisy digits. This is because both BAM and BPNN has noise tolerance capability. So, it can be concluded that the proposed algorithm (BAM+BPNN) exhibits a very fast and robust approach in recognition of Bangla digits.

In future, we have a plan to train our network with more samples to improve the accuracy rate. Because the recognition accuracy of a hybrid neural network increases with the increase of training accuracy.

Acknowledgment

We would like to express our deep gratitude to the anonymous reviewers of this paper. Their useful comments have played a significant role in improving the quality of this work.

References

[1] "Bengali language." [Online]. Available: http://en.wikipedia.org/wiki/Bengali_language

[2] V. S. P. U. Roy, K. and B. Chaudhuri, Eds., *A system towards Indian postal Automation*, 2004.

[3] Y. Wen, Y. Lu, and P. Shi, "Handwritten Bangla numeral recognition system and its application to postal automation," *Pattern Recognition*, vol. 40, no. 1, pp. 99–107, Jan. 2007. [Online]. Available: http://dx.doi.org/10.1016/j.patcog.2006.07.001

[4] D. Singh1 and B. S. Khehra, "Digit recognition system using back propagation neural network," *International Journal of Computer Science and Communication*, vol. 2, no. 1, pp. 197–205, January-June 2011.

[5] R. B. Ashoka H. N., Manjaiah D. H., "Feature extraction technique for neural network based pattern recognition," *International Journal on Computer Science and Engineering (IJCSE)*, vol. 4, no. 3, pp. 331–339, March 2012.

[6] C.-L. Liu and C. Y. Suen, "A new benchmark on the recognition of handwritten bangla and farsi numeral characters," *Pattern Recogn.*, vol. 42, no. 12, pp. 3287–3295, Dec. 2009. [Online]. Available: http://dx.doi.org/10.1016/j.patcog.2008.10.007

[7] M. M. Hoque and M. O. Karim, "Extraction of meaningful fuzzy features for recognizing of bangla handwritten numerals," in *Computer and Information Technology (ICCIT), 2006 9th International Conference on*, 2006, pp. 61–66.

[8] N. Das, B. Das, R. Sarkar, S. Basu, M. Kundu, and M. Nasipuri, "Handwritten Bangla Basic and Compound character recognition using MLP and SVM classifier," *ArXiv e-prints*, Feb. 2010.

[9] M. A. F. M. A. S. M. I. A. S. M. Mahbub Morshed, Md. Shahidul Islam, "Automatic sorting of mails by recognizing handwritten postal codes using neural network architectures," in *Proceedings of the ICCIT (International Conference on Computer and Information Technology)*, 2007.

[10] M. S. S. Shariful Hasan Shaikot, Fahim Kawsar, "Bangali digit recognition system using 3-layer backpropagation neural network," in *Computer and Information Technology (ICCIT), 2003 6th International Conference on*, 2003, pp. 327–331.

[11] H. G. E. Rumelhart, D. E. and R. J. Williams, *Learning internal representations by error propagation. In Parallel distributed processing: Explorations in the microstructure of cognition*. Bradford Books, Cambridge, MA, vol. 1.

[12] P. D. M. A.-A. B. Sanjit Kumar Saha, Md. Shamsuzzaman, Ed., *On Bangla Character Recognition*. Dhaka, Bangladesh: IEEE Computer Society Press., 23-25 December 2010, proceedings of 13th International Conference on Computer and Information Technology (ICCIT 2010).

[13] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.

[14] S.-W. L. Lam, L. and C. Y. Suen, "Thinning methodologies-a comprehensive survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 9, p. 879, September 1992.

A Distributed SCTP Scheme for Bandwidth Aggregation

Mohammad Mamun Elahi, Shahin Khan
Department of Computer Science and Engineering
United International University
Dhaka, Bangladesh.
mmelahi@cse.uiu.ac.bd, shahinkhanbsc@yahoo.com

***Abstract*—Bandwidth aggregation is a popular research topic that enables users to use different kind of network interfaces connected with Internet together. This is very useful especially in modern smart phones, where multiple interfaces can be used not only to aggregate bandwidth, but also to use an uninterrupted network connection. For different reasons, service providers may not be able to provide service all places according to customers' need. Also, Line may be dropped for mobility. Considering all those factors, we have developed an algorithm that extends basic Stream Control Transmission Protocol (SCTP) to aggregate bandwidths of multiple interfaces. In this paper, we have simulated the effects of different related parameters using C programming language and measured and demonstrated performance of the suggested algorithm graphically.**

***Keywords*—SCTP, multistreaming, multihoming, multipath data transfer, bandwidth aggregation.**

I. Introduction

TCP is a reliable protocol but slower. UDP is faster but not a reliable protocol. Also, TCP or UDP use single stream to transfer data. Using a single stream has a problem that, if data loss happens at any point in the stream then it blocks the delivery of the rest of the data. SCTP combines the features of TCP and UDP. SCTP allows multi-streaming service [2]. If one of the streams blocks, then it can use another stream. So, SCTP has less probability to drop line then that in the TCP. Basic SCTP cannot use more than one stream at a time. It utilizes other streams only for backup. So, bandwidth aggregation is not possible. Another scheme is CMT-SCTP (Concurrent Multipath Stream Transfer Protocol) [7]. It can use more than one stream at a time, but its goal is not to aggregate the bandwidths of different interfaces simultaneously.

In this paper, we have proposed an extended SCTP protocol, named as Distributed Stream Control Protocol (DSCTP) that supports multi-homing and multi-streaming to aggregate bandwidths of multiple interfaces of a device. DSCTP can use all the available active streams simultaneously at the same time. We have simulated and graphically demonstrated the efficacy of the approach.

The paper is organized as follows. Section II gives some of the related works. Section III describes the proposed scheme. Section IV analyzes the experimental results graphically. Section V describes some limitations and suggests some future works. Finally, section VI concludes the paper.

II. Related Works

Based on Resource Pooling (RP) ideas, in paper [1] the authors proposed a modified Resource pooling Congestion control for CMT-SCTP, which is denoted as CMT/RP-SCTP. In the paper, they combine Concurrent Multipath Transfer (CMT) with the idea of Resource Pooling (RP) in order to achieve a performance improvement over non-CMT transfer while still remaining fair to concurrent flows on congested links. Unlike existing approaches, which adapt classic TCP to a multi-homed CMT protocol; their approach does not depend on specific characteristics of TCP. Their approach is generic and may be adapted to other protocols quite as well.

The authors in [2] identify three negatives side effects of re-ordering introduced by CMT. Such as (i) Unnecessary fast retransmissions at the sender, (ii) Reduced congestion window (cwnd) growth due to fewer cwnd updates at the sender, and (iii) More ack traffic due to fewer delayed acks. They propose three algorithms which modified current SCTP algorithm and solve those problems.

In [3], a comprehensive review of transport layer multi-homing using the Stream Control Transport Protocol is provided. The main areas of study include handover management, concurrent multipath transfer and cross-layer activities. A survey of strategies among research efforts has brought improved performance while in some cases even complete solution.

The authors of paper [4] examined the challenges of CMT in asymmetric setups, which are very realistic for multi-homed Internet sites. They identified the problems of sender and receiver queue blocking, leading to poor performance of CMT. By applying their proposed mechanisms Buffer Splitting, Chunk Rescheduling and Smart Fast Retransmission, a significant performance improvement can be achieved – for plain CMT as well as for CMT combined with RP.

In [6], the authors identified the failure-induced receive buffer (rbuf) blocking problem in Concurrent Multipath Transfer (CMT) using SCTP multi-homing, and proposed CMT with a potentially-failed destination state (CMT-PF) to alleviate rbuf blocking. In the paper, they have shown

evaluation of CMT vs. CMT-PF. Using ns-2 simulations, they have shown that CMT-PF performs on par or better than CMT during more aggressive failure detection thresholds than recommended by RFC4960. They have also examined whether the modified sender behavior in CMT-PF degrades performance during non-failure scenarios. Their evaluations consider: (i) Realistic loss model with symmetric and asymmetric path loss, (ii) Varying path RTTs. They have found that CMT-PF performs as well as CMT during non-failure scenarios, and interestingly, outperforms CMT when the paths experience asymmetric rbuf blocking conditions. They recommend that CMT be replaced by CMT-PF in future CMT implementations.

From [7] we find, for technical reasons, SCTP uses one network path for data transmission and utilizes the other paths for backup only. An extension to support the load balancing of user data onto multiple paths in order to increase the payload throughput is Concurrent Multipath Transfer for SCTP, denoted as CMT-SCTP. In this paper, they described their CMT-SCTP extension for the SCTP model provided by the INET framework. Instead of using a single network path for user data transport and keeping the others for backup only, it is desirable to utilize all paths to improve the payload transmission throughput. An approach is denoted as CMT-SCTP, has been proposed by [8] and it is already implemented in the FreeBSD kernel. However, in order to examine open topics on the usage of CMT-SCTP for MPI (Message Passing Interface) or issues on its fairness against TCP have made a simulation model necessary.

The authors in [10] have investigated a novel approach that uses multiple network interfaces on cluster nodes to improve the performance of MPI programs. Unlike systems that support multi-railing in the middleware, their approach uses CMT, an extension to SCTP, which supports multi-railing at the transport layer in the kernel. CMT dynamically adapts to changing network conditions, and therefore simplifies the network configuration process. They showed that CMT is able take advantage of the additional bandwidth of an additional interface, and compares well with Open MPI's load balancing scheme. Experiments with MPI collectives demonstrate the possible advantages in scheduling messages at the transport layer (using CMT) versus scheduling messages in the middleware (using Open MPI). CMT shows promise for bandwidth intensive applications.

In [11], the authors have shown the difficulties of Multi-Path transport in a real-world dissimilar path setup which consists of a high speed fiber optic link and an ADSL connection. After that, they presented an optimized buffer handling technique which solves the transport efficiency issues in this setup. Their optimization is first analyzed by simulations. Finally, they also have shown the usefulness of their approach by experimental evaluation in a real Multi-Homed Internet setup. The authors in [12], have shown how CMT-SCTP data transport performance can be improved by combining multi-streaming with an advanced stream scheduling policy and SCTP API enhancements. Using CMT with a single byte stream is a challenging task, particularly when the memory footprint and management complexity have to be kept reasonably small. But unlike MPTCP, the SCTP protocol supports multi-streaming. Using CMT-SCTP for SCTP-based CMT transport, they have introduced an advanced stream scheduling policy to achieve an improved service performance: by mapping each stream to a certain path, the buffer sizes can be kept small. In comparison to round-robin scheduling, their approach utilizes the network with smaller buffer sizes while also reducing the per-stream message delay.

The authors of [13] demonstrate using simulation that if two paths are used for CMT, the lower quality path degrades overall throughput of an rbuf (Receive Buffer) constrained CMT association by blocking the rbuf. They argue that rbuf blocking is not specific to the transport layer, but applies to multipath transfers at other layers as well. They present and discuss CMT performance using several retransmission policies and various constrained rbuf values. They also study the impact of rbuf blocking with different combinations of end-to-end loss rate and delay on the two paths and show that when large differences exist in path delays and loss rates, using only the better path outperforms using two paths concurrently. While rbuf blocking cannot be eliminated, it can be reduced by choice of retransmission policy - a mechanism available to only the transport layer.

In [14], the authors have proposed a method that supports UMTS/WLAN vertical handover using SCTP, More specifically a dynamic address reconfiguration (DAR) extension called MSCTP (Mobile SCTP). The proposed method is useful for supporting vertical handover between any heterogeneous wireless networks in general and is not limited to UMTS and WLAN.

Most of the research on SCTP so far was based on basic SCTP, where regardless of how many streams are available, only a single stream is used at a time as primary path, and an alternate stream as backup path. In these category of schemes, though fault tolerance is better, parallel transmission to aggregate bandwidths of different interfaces is not possible. To utilize different bandwidths of multiple interfaces, many researchers already started to develop algorithms to employ SCTP with the capability of using multiple streams simultaneously. But there are lots of difficult problems to deal with, such as:

- distributing packets to different streams efficiently
- congestion control in all the streams
- selecting and dropping streams as necessary etc.

Considering all the factors from the literature review, we have suggested an algorithm to extend basic SCTP and we mainly focused on the distribution of data into different streams and show how that improves the performance.

III. PROPOSED SCHEME

A. Why DSCTP?

We are proposing a protocol called Distributed Stream Control Protocol (DSCTP) that supports multi-homing and multi-streaming. So it can use multiple interfaces at a time. The use of multiple interfaces simultaneously opens new way

of addressing some of the limitations of wireless media and can enable other new and interesting possibilities, such as:

- Bandwidth Aggregation
- Mobility Support
- Reliability
- Resource Sharing

B. Proposed Algorithm

There are two cases of DSCTP algorithm based on available active streams.

Case 1: If one Stream is active, then it works as Basic TCP.

Case 2: If n stream is active:

Suppose, Total data (File Size) = D_t

Cluster Size = C_s

Remainder = K

Total Stream = n

Then, the Cluster size C of cluster # 1 to (n-1)

$= (D_t - K)/n$

The Cluster size C of cluster # n

$= ((D_t - K)/n) + K$

Stream 1 will be used to send the data of Cluster 1 and same way, Stream n will be used to send the data of Cluster n. For example, suppose, File Size D_t = 512 kb, Remainder K = 2, and Total Stream n = 3. Then,

Cluster size of Cluster # *1 to (n-1)* = (512-2)/3 = 170

Cluster size of Cluster # *n* = ((512-2)/3) + 2 = 172

For our experiment, UMTS will be used to send the data of Cluster 1, WLAN will be used to send the data of Cluster 2 and Cable Line will be used to send the data of Cluster n.

C. Dropped Stream and Free Stream

If any stream drops for any reason, then the remaining packets scheduled for dropped stream will use the stream, which will be available first. If there is more than one free stream, then the algorithm will use all free streams and divide data according to case 2. Connection manager will do all that and will monitor all the activities of entire stream. If any stream completes data transmission and becomes idle, then remaining packets scheduled to other streams will be re-calculated and will be re-scheduled to use the free streams in parallel. Here, case 2 from above algorithm will be used.

D. How to decide a stream is drop

Once a data chunk is reported as missing for three times, it is retransmitted immediately on the same stream. If retransmission fails again, it will use alternative (next available) stream. If same thing happens three times consecutively, then it will consider the stream as dropped. Connection manager will transmit remaining data using alternative streams. Also, connection manager will monitor the dropped stream using Heart Bit for a certain time [5]. If the stream gets reactivated by this time, it will be used again.

E. Hand off

If any stream's performance is poor or not as expected, it continues working until another stream is available or free. After another stream becomes free or available, a handoff will occur and the previous stream will be used as a backup stream only [14].

F. Assumption

We have assumed for our experiment that, bandwidth for all streams is equal (16 Kilo Bits/Second). We transfer data between 02, 06, 12, 18 and 20 streams using SCTP and DSCTP protocol.

IV. Result Analysis

We have simulated and have shown the comparative results among TCP, SCTP and DSCTP using the following parameters:

- Transfer Time
- Data Loss
- Throughput
- Number of Retransmission

A. Calculation the transfer time

We have used the following equation to find out the transfer time of selected data using TCP, SCTP, and DSCTP Protocol.

$$\textit{Data Transfer Time} = MAX\,(S_0\ time: S_n\ time) + Cmd + Rd + Cmd + Rd$$

Where, Cmd = Connection Managers Delay

Rd = Time for Retransmitted data

S_0 = Stream *0*

S_n = Stream *n*

TABLE I. Transfer Time

Original Data Size	Transfer Time		
	TCP	SCTP	DSCTP
512	52.44	52.14	23.26
1024	104.58	97.86	46.70
2048	204.02	193.88	96.02
4096	402.59	400.29	192.51
8192	804.05	789.51	420.56
16384	1604.86	1578.93	798.35

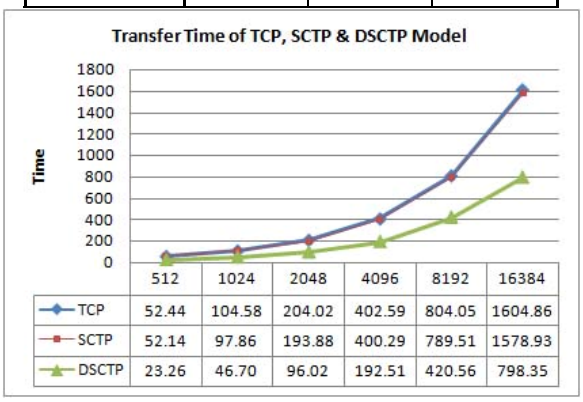


Fig. 1. Comparison of transfer time

B. Calculation the Data Loss

We use the following equation to find out the data loss of selected data using TCP, SCTP, and DSCTP protocols.

Data Loss = (DS - TDS)

Where, DS = Data Size (Original)

ToR = Time of Retransmission

TABLE II. DATA LOSS

Original Data Size	Data Loss		
	TCP	SCTP	DSCTP
512	48	32	0
1024	80	64	0
2048	160	32	16
4096	240	208	32
8192	480	240	336
16384	880	464	336

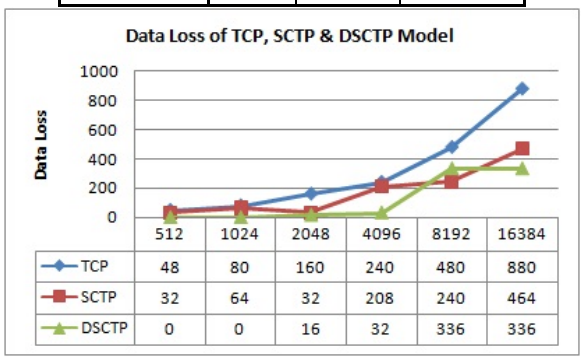


Fig. 2. Comparison of data loss

C. Calculation the Throughput

The following equation is used to find out the throughput of TCP, SCTP, and DSCTP.

Throughput = TDS / TT

Where, TDS = Transmitted data Size

TT = Transfer time

TABLE III. THROUGHPUT

Original Data Size	Throughput		
	TCP	SCTP	DSCTP
512	8.85	9.21	22.01
1024	9.03	9.81	21.93
2048	9.25	10.4	21.16
4096	9.58	9.71	21.11
8192	9.59	10.07	18.68
16384	9.66	10.08	20.1

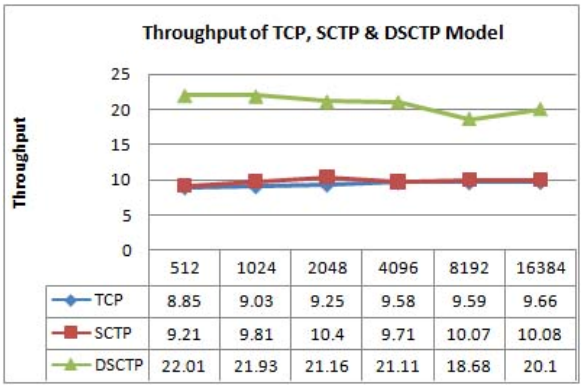


Fig. 3. Comparison of throughput

D. Number of Retransmissions

We show here the time of retransmission of selected data using TCP, SCTP, and DSCTP protocol.

TABLE IV. RETRANSMISSION TIME

Original Data Size	Retransmission Time		
	TCP	SCTP	DSCTP
512	3	2	0
1024	5	4	0
2048	10	2	1
4096	15	13	2
8192	30	15	21
16384	55	29	21

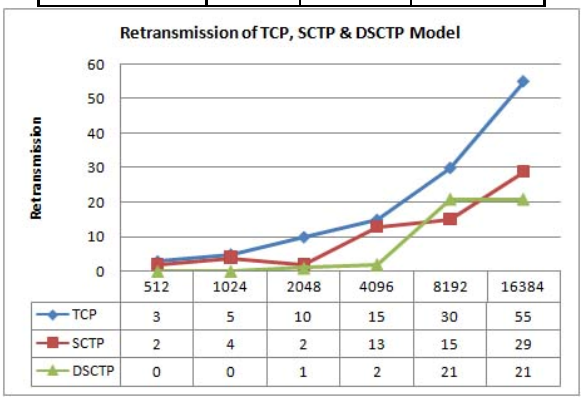


Fig. 4. Comparison of retransmission time

E. Overall Comparison

We show here a comparison between 02, 06, 12, 18 and 20 streams using DSCTP for Transfer Time, Data Loss, Throughput, and Number of Retransmission. Comparison of transfer time of multiple paths using DSCTP:

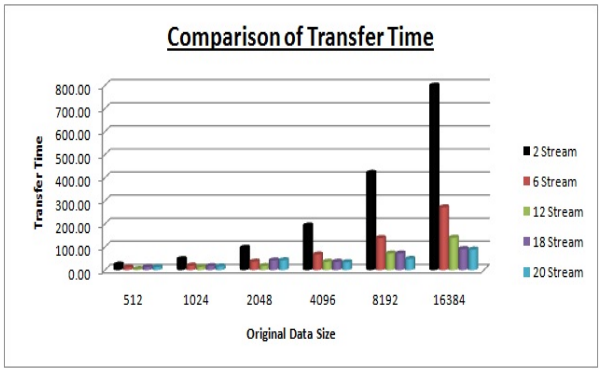


Fig. 5. Comparison of transfer time

TABLE V. TRANSFER TIME

Original Data Size	Transfer Time				
	2 Stream	6 Stream	12 Stream	18 Stream	20 Stream
512	23.26	10.96	4.14	12.77	12.89
1024	46.70	19.06	13.77	16.97	15.42
2048	96.02	36.35	17.42	41.57	42.00
4096	192.51	66.95	35.68	35.49	33.53
8192	420.56	138.40	71.31	71.62	47.66
16384	798.35	271.38	139.56	90.75	88.26

Comparison of data loss of multiple paths using DSCTP:

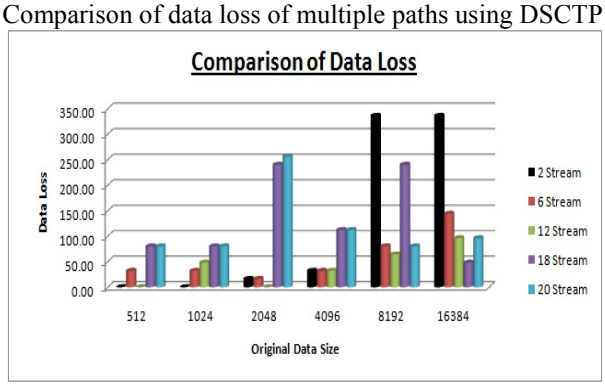


Fig. 6. Comparison of data loss

TABLE VI. DATA LOSS

Original Data Size	Data Loss				
	2 Stream	6 Stream	12 Stream	18 Stream	20 Stream
512	0.00	32.00	0.00	80.00	80.00
1024	0.00	32.00	48.00	80.00	80.00
2048	16.00	16.00	0.00	240.00	256.00
4096	32.00	32.00	32.00	112.00	112.00
8192	336.00	80.00	64.00	240.00	80.00
16384	336.00	144.00	96.00	48.00	96.00

Comparison of throughput of multiple paths using DSCTP:

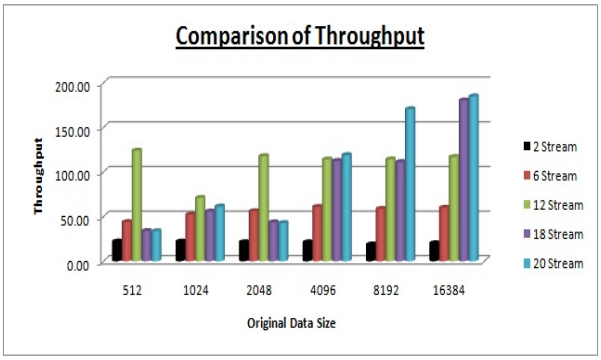


Fig. 7. Comparison of throughput

TABLE VII. THROUGHPUT

Original Data Size	Throughput				
	2 Stream	6 Stream	12 Stream	18 Stream	20 Stream
512	22.01	43.80	123.76	33.82	33.52
1024	21.93	52.04	70.90	55.62	61.20
2048	21.16	55.91	117.57	43.49	42.67
4096	21.11	60.70	113.92	112.26	118.81
8192	18.68	58.61	113.99	111.03	170.22
16384	20.10	59.84	116.71	180.00	184.55

Comparison of number of retransmission of multiple paths using DSCTP:

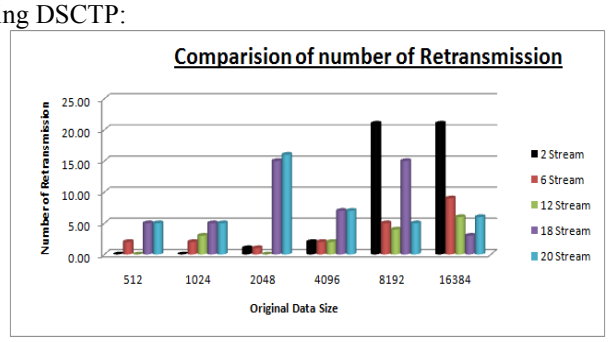


Fig. 8. Comparison of number of retransmission

TABLE VIII. NUMBER OF RETRANSMISSIONS

Original Data Size	Number of Retransmission				
	2 Stream	6 Stream	12 Stream	18 Stream	20 Stream
512	0.00	2.00	0.00	5.00	5.00
1024	0.00	2.00	3.00	5.00	5.00
2048	1.00	1.00	0.00	15.00	16.00
4096	2.00	2.00	2.00	7.00	7.00
8192	21.00	5.00	4.00	15.00	5.00
16384	21.00	9.00	6.00	3.00	6.00

F. Reason for good output

We show the result for 2, 6, 12, 18 and 20 Stream using TCP, SCTP and DSCTP protocol. SCTP use 01 Stream at a

time. So, the output of SCTP is not as expected if we compare it with TCP. But the probability of line drop is less than TCP. On the other hand, DSCTP use all streams at the same time. So the output is 50.25% - 55.65% better than TCP and 50.56% - 55.39% better than SCTP (for 02 streams, according to Transfer time). Also, the output is 66.76% - 67.39% better than TCP and 66.22% - 67.20% better than SCTP (for 03 streams, according to Transfer time). If we use more streams, then the result will be better until a certain level. It will depend on data size and bandwidth.

V. Limitations and Future Works

We did not decide that maximum how many paths will provide us the best result. We also have assumed that all paths' bandwidth is same. If the bandwidths differ, our algorithm should work properly but the throughput will decrease and data overhead will increase. Because, we need to maintain more 'Free Paths' and 'Hand Offs' to get more realistic results. We have proposed a way to maintain 'Free Paths', 'Dropped Paths' and 'Hand Offs'. But there are still more scopes to develop the algorithm.

DSCTP is our ongoing project. We are currently examining its performance more specifically. We are working to solve different kinds of technical problems and working to find out how many paths will provide us the best performance. Also we are working to solve the problem of maintaining 'Free Paths', 'Dropped Paths' and 'Hand Offs'.

VI. Conclusions

It is very important for modern devices to use all available network interfaces at a time. It will improve its data transfer time as well as throughput. We have developed an algorithm to aggregate bandwidth of different interfaces and have analyzed and shown that our scheme provides better performance than TCP and basic SCTP. If number of available streams increases, performance gets better, but maintenance overhead also increases. So optimal number of streams needs to be calculated based on total data size.

References

[1] T. Dreibholz, M. Becke, J. Pulinthanath, and E. P. Rathgeb, "Applying TCP-Friendly Congestion Control to Concurrent Multipath Transfer," In *Proceedings of the IEEE 24rd International Conference on Advanced Information Networking and Applications (AINA)*, Perth/Australia, April 2010.

[2] J. Iyengar, K. Shah, P. Amer, and R. Stewart, "Concurrent Multipath Transfer Using SCTP Multihoming," in *SPECTS*, California, July 2004.

[3] T. D. Wallace and A. Shami. A Review of Multihoming Issues Using the Stream Control Transmission Protocol. Department of Electrical and Computer Engineering, The University of western Ontario, Canada, 13 June 2011.

[4] T. Dreibholz, M. Becke, E. P. Rathgeb, and M. T¨uxen, "On the Use of Concurrent Multipath Transfer over Asymmetric Paths," in *Proceedings of the IEEE Global Communications Conference (GLOBECOM)*, Miami, Florida, USA, Dec. 2010.

[5] B. Forouzan and S.C. Fegan. Data Communications and Networking. McGraw Hill, 2006.

[6] P. Natarajan, N. Ekiz, P. D. Amer, J. R. Iyengar and R. Stewart, "Concurrent Multipath Transfer using SCTP Multihoming: Introducing the Potentially-failed Destination State," 2008.

[7] T. Dreibholz, M. Becke, J. Pulinthanath, and E. Rathgeb, "Implementation and Evaluation of Concurrent Multipath Transfer for SCTP in the INET Framework," In *Proceedings of the 3rd ACM/ICST International Workshop on OMNeT++*, Malaga, Spain, March 2010.

[8] J. R. Iyengar, P. D. Amer, and R. Stewart, "Concurrent Multipath Transfer Using SCTP Multihoming Over Independent End-to-End Paths," *IEEE/ACM Transactions on Networking*, 14(5):951-964, 2006.

[9] T. Dreibholz, M. Becke, H. Adhari, E. P. Rathgeb. On the Impact of Congestion Control for Concurrent Multipath Transfer on the Transport Layer. University of Duisburg-Essen, Institute for Experimental Mathematics Ellernstraße 29, 45326 Essen, Germany.

[10] B. Peno, M. Tsai, J. Iyengar, and A. Wagner, "Using CMT in SCTP-based MPI to Exploit Multiple Interfaces in Cluster Nodes," in *Proceedings of the EuroPVM/MPI*, Paris/France, September 2007.

[11] H. Adhari, T. Dreibholz, M. Becke, E. P. Rathgeb, and M. T¨uxen, "Evaluation of Concurrent Multipath Transfer over Dissimilar Paths," in *Proceedings of the 1st International Workshop on Protocols and Applications with Multi-Homing Support (PAMS)*, Singapore, March 2011, pp. 708–714.

[12] T. Dreibholz, R. Seggelmann, M. T¨uxen, and E. P. Rathgeb, "Transmission Scheduling Optimizations for Concurrent Multipath Transfer," in *Proceedings of the 8th International Workshop on Protocols for Future, Large-Scale and Diverse Network Transports (PFLDNeT)*, vol. 8, Lancaster, Pennsylvania, USA, Nov. 2010, ISSN 2074-5168.

[13] J. R. Iyengar, P. D. Amer, and R. Stewart, "Performance implications of a bounded receive buffer in concurrent multipath transfer," *Comput. Commun.*, vol. 30, no. 4, pp. 818–829, 2007.

[14] L. Ma, F. Yu, and V. C.M. Leung. A new method to support UMTS/WLAN Vertical Handover using TCP. The University of British Columbia, Tejinder Randhawa, British Columbia Institute of Technology, 24 August 2004.

University Course Advising:

Overcoming the challenges using decision support system

Tanvir Roushan, Dipankar Chaki, Onishim Hasdak, Md. Syeed Chowdhury, Annajiat Alim Rasel, M. Abdur Rahman, Hossain Arif
Department of Computer Science and Engineering
BRAC University
Dhaka, Bangladesh
Email: tanvir.roushan@gmail.com, joy.dcj@gmail.com, onishim.hasdak@gmail.com, grnsyeed@gmail.com, annajiat@gmail.com, anon83@gmail.com, hossain.arif@gmail.com

***Abstract*— Universities require a fast and reliable system to provide academic advising to its students, register them into different courses, and to manage change requests. Currently in Bangladesh, academic advising is done mostly on paper, as seen in many of the public and private universities. In this paper, we propose a unique online-based system that would make the university course advising and registration process easier and intelligent. The proposed online tool will work very closely and interdependently with each educational institution's own course curriculum, grading system, and university policies. The intention of this paper is to find key challenges in the system and to evaluate the efficiency of an easy-to-read web-based course advising and decision support tool. The tool expedites the academic advising process. It acts as an online advisor for the students, eases the task and save time of the advisors by automation of redundant tasks and reducing paper work. The recommended advising software through its decision support system aids both students and teachers with unique facilities. It also provides additional time for advisors to focus on students' development.**

***Index Terms*— online academic advising, course registration automation, decision support system**

I. Introduction

At the beginning of each semester of a university, every student has to register for a number of courses. This is called the course advising and registration. Course registration is not just selecting a course for an upcoming semester, from a pre-fixed schedule. This is an intricate and elaborate task for the faculty advisors and the students to sit together to select the course sequence throughout the academic life that would be most appropriate for a student. This requires planning and preparation that vary almost in every individual case. So advising is an experience that every student has to go through in order to peruse a successful academic life. Universities around the globe share the two systems of offering courses to students; one being where students have the liberty to choose courses of their choice, and the other is where the university fixes a set of courses throughout the program. The practice of both these systems subsists in Bangladesh. Many universities use software systems for course advising or web-based applications to register for a course. Administrative work is shifted from paper to computers [1]. Although application of IT has made the task of student management easier; in most cases, automated course advising is absent and application of technology is limited up to the registration for the courses for a semester. However, the student counseling still requires one to one session between advisor and the advisee.

Generally, the course advising and registration takes place on one day or two for a whole institution (up to a week for some institutions). This creates long queues in front of the advisor's office and the eventually floods the school. Students fail to take necessary counsel; they miss the opportunity to take time to speak with an advisor about courses, degree plans. Advisors are under tremendous pressure running under a tight schedule, let alone technical complications for example server overload, internet connectivity failure, printer jam etc. Most of the advisors' time is spent on repetitive tasks instead of assisting the students with efficient advice on reaching full academic potential during the educational experience in the institution [2]. Introducing technology to the advising process aims at leveraging repetitive tasks on software and dedicating time to helping a student plan his/her education roadmap. This paper proposes a system that aims to overcome the hassles faced before and during course advising, and which assists students in taking decisions with the help of a decision support system (DSS). The course advising, registration, and decision support system is web-based and interactive. A web-based application is any application that uses a web browser as a client. The software application is coded in a browser-supported programming language. The advisor and the advisee login to the application with credentials to view grade sheets, advising form, courses offered, schedule and so on. The system has all the academic rules built-in as well as a complete faculty, student, and course data. Both the student and the advisor can access this information quickly and easily.

II. Background And Challenges

Let us look deep into the scenario to identify the need of this system in the first place. Most of the universities have shifted to the use of technology in replacing paper work for student advising, in order to save time and focus more on the students themselves. Academic credit system is widely used in many educational organizations. As mentioned by Chau and Phung, knowledge discovery from educational data in a

semester credit system is an unsolved problem. To tackle those problems, the authors propose a knowledge-driven decision support that is helpful for educational managers to make appropriate and reasonable decisions about student's study [3]. For instance, examining the existing state of the information system in use at universities in Dhaka, we observed that the existing advising systems are not fulfilling the need of the advisor or the advisee, completely. There are scopes for improvement beyond regular tasks. An advisor according to Heisserer and Parette is, higher education professionals who understand students' challenges. They are the primary candidates for advisor/mentor roles [4]. While faculty member, administrators, and students' affair professionals all serve as student advocates and play an integral part in student retention, advisors are typically in the best positions to assist students in making quality academic decisions. In addition, an advisor should be knowledgeable about academic programs and curricula requirements within the institution. Students are expected to be given accurate and correct academic guidance [5]. Through the process of advising, the students decide which courses they register for and the advisee is responsible to supervise if the students take the correct and a valid decisions. The advised courses should uphold the university's academic policies and avoid schedule overlapping.

Pre-advising is a computerized system of telling the university about student's personal details and course intentions, followed by the tangible advising. Pre-advisement gives the student the opportunity to avoid long waiting queues during heavy registration, by taking time to speak with an advisor about courses, degree plans and transferring [6]. Students fill out their registration form, with course information beforehand, so that their registration day goes smoothly. It is necessary because there are loads of students who are willing to register a course at the same time. Universities can only allow a certain number of students to be enrolled in each course. The arrangement of pre-advising reduces pressure during advising at the end of the semester. This pre-advising step is absent in most universities except very few. Usually course registration and advising is carried on right after the result is published. Some limitations exist in the current pre-advising and advising system. Such as the lack of sufficient time available for students to register for a course. Often collective information about the course pre-requisites, class and examination schedule are not available to the students. The small window of time that a student gets while advising is insufficient. There are both software issues and hardware limitations. There is a huge time gap between the pre advising and advising. Students need to wait more than a month after pre-advising to finalize their advising for the next semester. Therefore, a student might have to change his/her section or add/drop course depending on the grades earned at that semester among various other issues. The students cannot amend their list of courses selected during pre-advising until the actual advising is done and unless an advisor approves them to make the change.

The system has a way to control priority that allows senior students to complete pre-advising in advance. If a student misses the particular time allocated for him/her to register for a course, they may have to go through a lot of trouble, which may include server overload, system breakdown, session time out etc. As a result, the student might miss the opportunity to reserve a seat of a course. This can create longer sufferings later on. During advising, one has to be present physically at the university to meet the advisor or drop slips and complete the advising process. Both students and teachers face difficulty while going through the advising process under the current system. Limited server availability, inadequate and slow response time, server overload are common. These problems eventually lead to server failure resulting to system halt. Students are deprived from their thorough counseling due to a short session with the advisor alongside technical difficulties and human resource inefficiency.

Technology should be used to improve academic services and facilitate the access of information to the students. However, this should not replace the face-to-face student-advisor interactions [7]. An automated advisor gives answers to the student's routine questions. The student can then meet with the advisor for further consultation. This combination of human and machine can save time for the human advisor [8]. We aim to address the challenges that the faculty members and students of universities from all over Bangladesh face during the course advising and registration every semester. Based on these issues, the need of a new intermediary system was felt for a long time. This paper discusses a decision support system (DSS) based IT solution to resolve the difficulties that the current advising systems have. The course advising and decision support system software will benefit the students, by overcoming the limitations of any current advising method used in universities all over Bangladesh. The web based DSS software ease the process, reduces the overall time needed to do the advising, and is able to fulfill the actual purpose of advising.

III. System Architecture

Students of the university have to complete their pre advising and advising each semester to take courses for the upcoming semester. The decision support system assists students in choosing the best set of courses in a semester. To be able to do that, the system must have access to the students' personal data and grade sheets, courses being offered, their schedule and available seat status etc. Gaining access to these data is not straightforward. Usually sensitive issues like adding and dropping a course is done under the supervision of the registrar's office of a university. This is where this advising tool comes in handy. All the data that are required are available in the system. An approach described by Mansmann and Scholl is, the decision support system integrates input data from relevant sources into an autonomous data warehouse [9].

The whole software system acts as an intermediary model that links the pre-advising and the advising, the registrar's office and the students. The illustration given below in Fig. 1 portraits the idea, how the existing advising system is bridged with the intermediary software to overcome the challenges. The decision support tool and course advising system includes

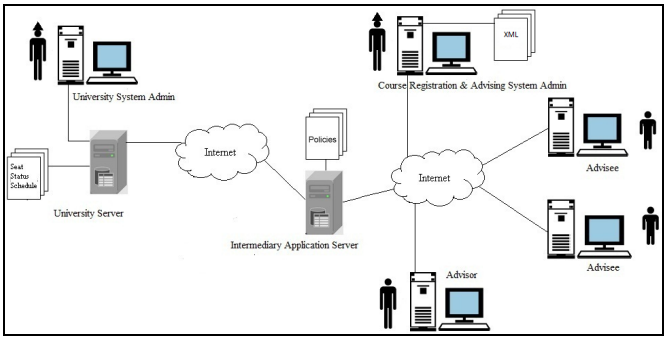


Fig. 1. System Architecture of the Online Advising Tool

all the required data; such as the student list, student grade sheet, university policies, advising rules etc. The class and examination schedule, seat status, course list offered and the ones not offered are fetched from the university website or from the registrar's office in absence of a website.

The proposed online advising tool has three nodes from the user point of view. The user of this system comprises of the advisee, the advisor, and the administrator. The advisor conducts the course advising with the intermediary application server. The automated software allows students to select courses they desire to their advising form, however registering for those course must be valid under the policies and regulations of the particular educational institution. These rules and guidelines are handled automatically by the decision support system, and the advisor does not need to be bothered about students taking the correct and valid course. From the software engineering point of view, the system is developed by following mostly the waterfall methodology approach. The waterfall model is a sequential design process, in which progress is seen as flowing steadily downwards through the phase of requirements definition, system design, software design, programming, testing, integration, operation, and maintenance [10]. Although waterfall model approach is a linear and one-pass process, the system prototype has also seen incremental development rapid and flexible response to changes. Figure 2 shows the use case diagram of the course registration and the decision support system. It is the simplest representation of the user's interaction with the system. Three actors of the system are shown here.

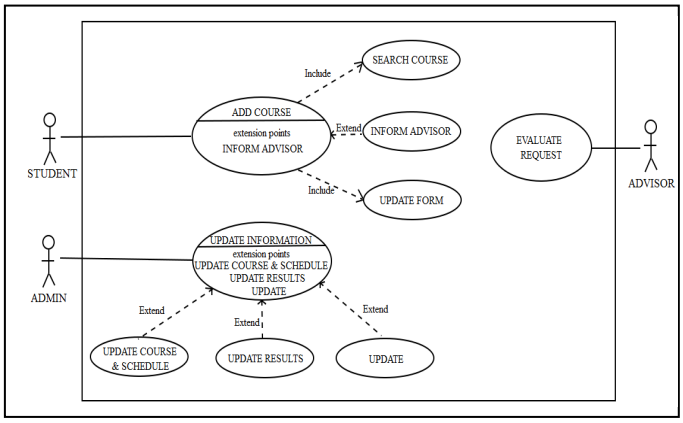


Fig. 2. User case diagram of online advising tool

The students register for courses with the assistance of the decision support system, administration interacts with the system making necessary changes in the system and the advisor evaluate the requests made by the advisee.

IV. Automation In Academic Counseling

The use of technology in academic advising may introduce greater accountability and may provide better services to students. However, technology does not replace one-on-one interactions. A few mentionable questions answered by the system itself are, i) if a course has a pre-requisite or an equivalent course, shown in Fig. 3 ii) if a student is allowed to register for courses above a limit based on the current cumulated grade point average (CGPA), shown in Fig. 4. In figure 3, it shows a student who tried to register for a course EEE207. The decision support system responds by showing the incomplete pre-requisite courses required to register for the course. In this particular case according to the university course registration policy, the system shows that the student must complete two courses EEE205 and EEE206. Figure 4 below shows that the system is not allowing a student to register more than four courses at a time, and showing a course registration exception. If any university course registration regulations state that their students will only be allowed to take five courses in a semester only if the student is successful in maintaining a constant CGPA over 3.5 on a scale of 4.0, then the system will be needed to be altered likewise. The automatic advising tool can be tuned according to the needs of an institutions.

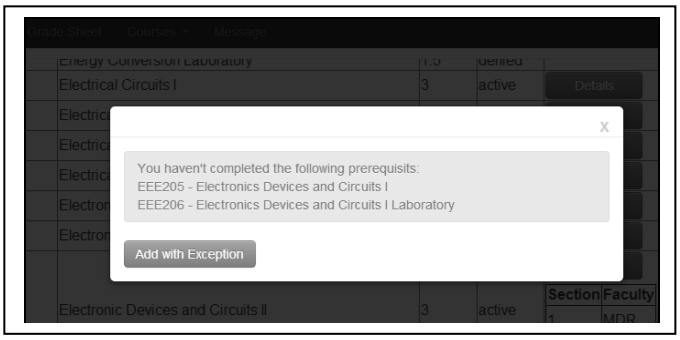


Fig. 3. Course Pre-requisite Alert Message

Fig. 4. Course Add Exception

A course registration might need academic advising, review or approval from the advisor. The inbuilt messaging facility of the proposed system let the students and teachers communicate among themselves regarding any issues that they might have regarding registering a course (Figure 5). An advisee might have an exceptional case that could possibly require the attention of the advisor. Often there are situations when the system does not allow students to register for a course. However if the student is confident enough, they always have the option to reach the advisor by adding a course with exception (Figure 3). Students login to the software and have the privilege to view their grade sheets, class and examination schedule, course list, both that are offered and the ones that are not. The seat status is a key element for the decision support system to take/provide a decision. Since the seat booking record is an official data it has to be periodically updated from an authorized source by the administrator in charge of this intermediary system.

The students and faculties are aided with the advising form. It shows the routine of all the courses that a student has registered. It also shows the examination routine of those courses. Clashes between the course schedule are shown immediately when a selected course overlaps with the timing of another course that is already selected in the advising form. Students can easily see the available courses and the sections offered, then select the routine that is most suitable. They can check the course details, like the faculty who is taking the course, the number of sections offered. One can also know the associated labs hours of a course and its routine before selecting a course.

Figure 5 shows a screenshot of the messaging window. The advisees can write a messages describing their problem, and explain the situation. The advisors can reply to the students by taking appropriate steps to solve that issue. Otherwise, a teacher may call upon a student to the office to deal with a problem face to face. One of the unique features of this online advising tool is the option of polling for a course (Fig. 6). The students can access the course list of the university. If required, a student can poll for a course that is not offered in the upcoming semester. Once poll in open anyone can take part in that poll.

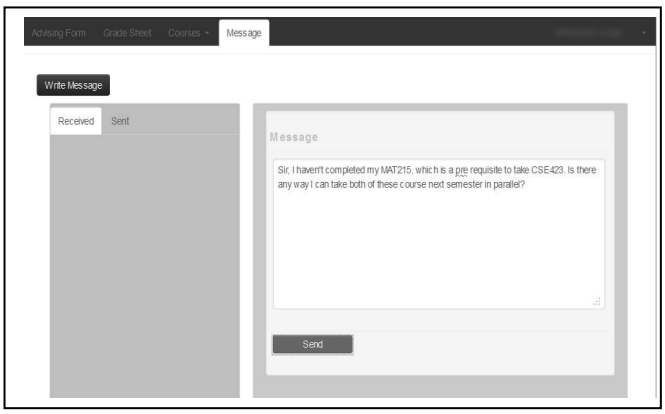


Fig. 5. Message sent among advisors and the students

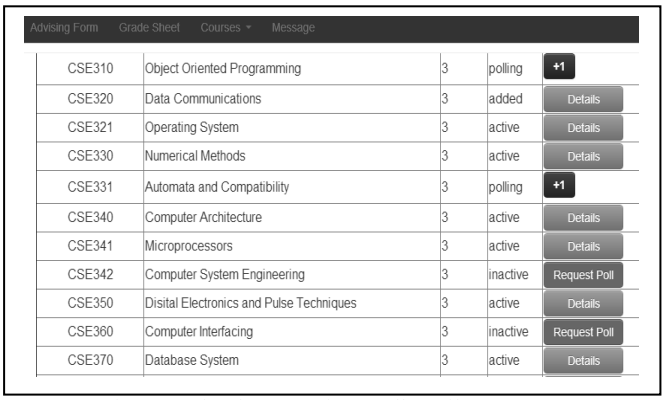


Fig. 6. Option for requesting a poll to offer a course

An admin has the liberty to allow or deny a poll request. If enough vote is accumulated for a particular course through a poll, the university authority might rethink about offering that course for the next semester based upon student demand. This reduces confusion among students and saves a lot of time of the university authority. The courses that are not offered are listed. The option to request a poll is available as shown in Fig. 6 above. In addition, all of these events are reported to the students and faculties. Everyone in the system is notified about the changes in the course advising system via messaging e.g., e-mail, in-built message inbox etc. Besides, the administrator can always broadcast any message to the users at any time. On the day of the advising, changing the courses repeatedly increases load on the university web server. Moreover, all of the students come to register for their course on that very day. Option for printing the grade sheet and advising form are available. Advisees can go to their advisors taking a hardcopy of their advising form on the scheduled day of their advising. Finally, students register for the courses in the university website or any paper system that subsists in the university. This reduces hassle and gives the teachers some time to communicate with the students.

V. The Online Advisor: A Decision support System

There are many factors and aspects to consider, reasoning to be done before one chooses for courses throughout his or her academic years. What if there was a system that would make these decisions for the students? Moreover, if the decisions are made based on intellectual analysis and in a rational manner! The web based decision support system automates this complex process of decision-making, and provides the students a handy assistance in choosing the academic course. The decision tree of the system is shown in Fig. 7. It portraits the major subject matter on which the decision system works. Gradually, the system makes more appropriate and reasonable decisions about student's education profile. The system initially prioritizes the courses based on the course IDs, assuming that courses numbers starting with ones (1) are first year courses, the course IDs that start with two (2) are second year course and so on. Students typically follow this sequence of courses in their academic life, and so does most of the university curriculum. Once all the courses are ordered accordingly, they

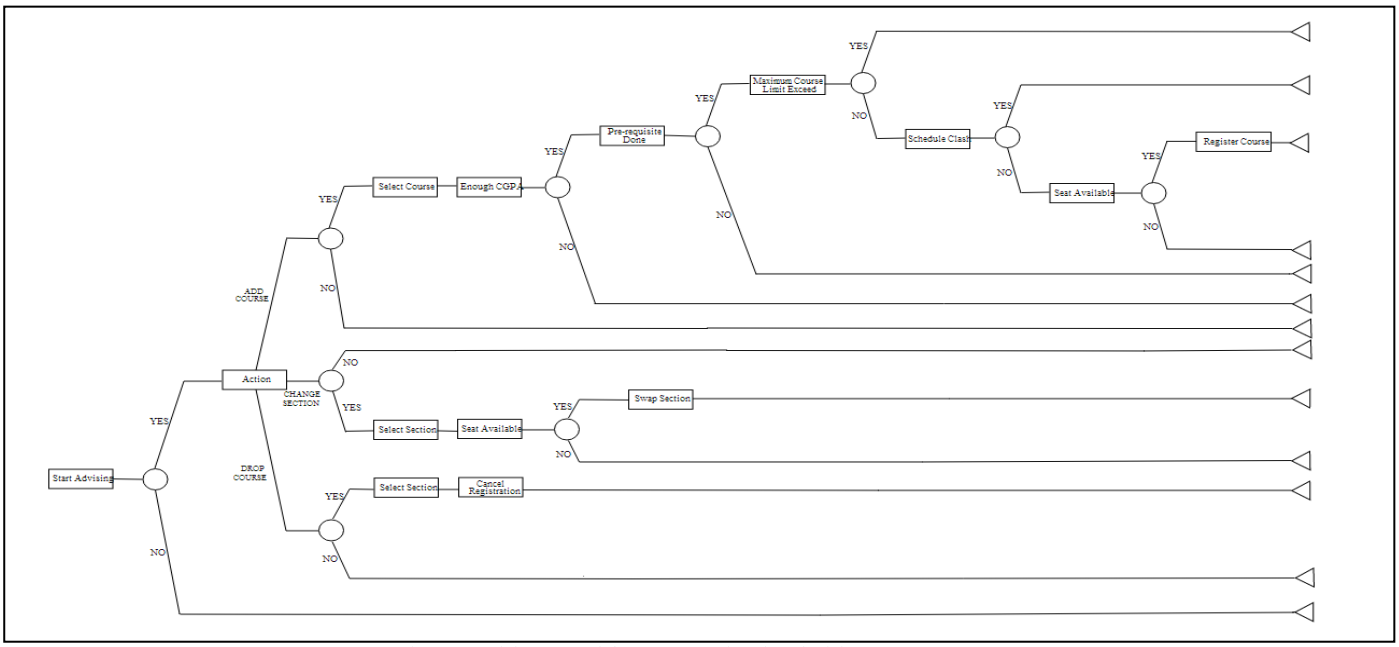
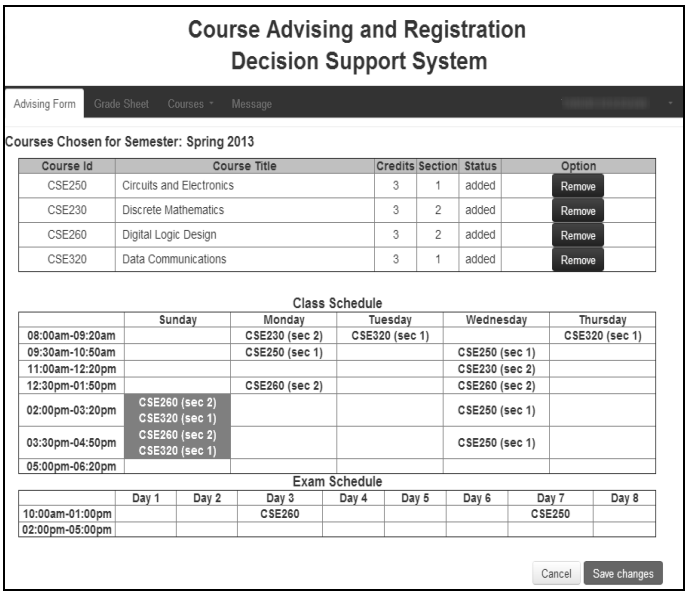


Fig. 7. Decision tree of the course registraion decision support system

are crossed checked with the grade sheets of the students. The grade sheet provides the information, which courses are already taken by a student. Then by eliminating the courses taken and seeking the courses that are higher in the priority queue according to the order made earlier, a list of available courses are recommended for each student. The accuracy of this process is increased by finding a set of courses that have non-conflicting class and examination schedule. Applying this naive yet intelligent algorithm the decision support system can narrow down the choices of courses that a student has. The performance of this process is improved by the addition of the non-department courses that are taken by students. For instance offering a humanities or a language course is very common among the universities of Bangladesh.



Course Advising and Registration Decision Support System

Advising Form | Grade Sheet | Courses | Message

Courses Chosen for Semester: Spring 2013

Course Id	Course Title	Credits	Section	Status	Option
CSE250	Circuits and Electronics	3	1	added	Remove
CSE230	Discrete Mathematics	3	2	added	Remove
CSE260	Digital Logic Design	3	2	added	Remove
CSE320	Data Communications	3	1	added	Remove

Class Schedule

	Sunday	Monday	Tuesday	Wednesday	Thursday
08:00am-09:20am		CSE230 (sec 2)	CSE320 (sec 1)		CSE320 (sec 1)
09:30am-10:50am		CSE250 (sec 1)		CSE250 (sec 1)	
11:00am-12:20pm				CSE230 (sec 2)	
12:30pm-01:50pm		CSE260 (sec 2)		CSE260 (sec 2)	
02:00pm-03:20pm	CSE260 (sec 2) CSE320 (sec 1)			CSE250 (sec 1)	
03:30pm-04:50pm	CSE260 (sec 2) CSE320 (sec 1)			CSE250 (sec 1)	
05:00pm-06:20pm					

Exam Schedule

	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7	Day 8
10:00am-01:00pm			CSE260				CSE250	
02:00pm-05:00pm								

Cancel | Save changes

Fig.8. Advising form showing a clashing class schedule

Figure 8 demonstrates a class schedule conflict situation where a student registered for four courses CSE230, CSE250, CSE260, and CSE320. However, there is a conflict between the two courses CSE111 and CSE220. The system shows an error on screen. Moreover it suggests a number of other courses that are available and suitable for that student.

Another exclusive feature of the advising tool is on probation warning. Students who are suffering to maintain a minimum level of CGPA are warned about their potential risk of going under probation. When they login to the system the warning is displayed. While advising, students are offered a guideline on how to get out from probation or not fall into one. Taking the CGPA into account and the past history of the student profile the system recommends a reduced number of courses.

The course list, class, and examination schedule, faculty list of spring 2013 of the same university was used as a test case. A class diagram for this is shown in Figure 3. The class and examination schedule, grade sheets were provided in XML format (xlsx, Microsoft Excel Spreadsheet). The database was designed in a manner so that the whole system is scalable; the search option within the system is made extremely fast to provide best user experience. The magnificent user interface is easy-to-read. No complex thought process is needed to use the system. Advance object oriented programming concept was practiced while developing the software. Rigorous testing was performed to find out the glitches within the system.

VI. Conclusion

The necessity of a comprehensive course advising system for the universities of Bangladesh carried this initiative of developing the online advising tool. Although no application software can be entirely fail proof, we developed a system that functions according to the need of the advisor and the advisee at a remarkable standard. The objective of this advising tool is

to make the current university course advising an improved and time efficient process. Using the decision support system can make students' experience of course advising more contented, alongside make advisors' task easier allowing both the advisor and advisee to link better and get the best outcome in the academic career. This system is quick and dependable, reduces the paper work, eliminated the redundant tasks, and automates the whole process of academic advising. An efficient and easy-to-read web-based course advising and decision support tool advances the academic advising process. However, counseling has to support the face-to-face meeting between advisor and student and not replace it except for optional prescriptive functions. It will allow advisors more time to focus on student development and contribute in making better planning and scheduling. Machines cannot replace humans yet, however this online advising tool takes us one-step closer to making academic advising exclusively programmed, provided enough information is available to the system. The worth mentioning decision support system, bridges the gap between the advisor and advisee, and aids the students in developing a clear road map of their academic career.

VII. Future Work

The success of the decision support system completely depends on the user satisfaction. A survey is planned to be conducted on the application software to get some statistics on which researchers and development team members can further find places to improve. An additional to-do for the near future may be to introduce a cellular phone gateway and integrate it with the system, so that the end users can be up to date with the notices and changes in the system via cell phone messages, rather than just relying to e-mail notifications.

The application software for university course advising and decision support can be further improved on its functionality and fresh features can be implemented. Using data mining techniques, this system may further be extended to assist in better academic scheduling, and predicting course demands.

Acknowledgment

We acknowledge the support of Professor M. Zahidur Rahman, Chairperson, and Assistant. Professor Ms. Sadia Hamid Kazi, Department of Computer Science and Engineering for their idea, inspiration, and overall suggestions to improve this work.

References

[1] G. Steele, M. Leonard, C. Haberle, and W. Lipschultz, "Technology and Academic Advising," Academic Advising News, 21, 2000.

[2] Al-Nory, M.T., "Simple Decision Support Tool for university academic advising," Information Technology in Medicine and Education (ITME), 2012 International Symposium, vol.1, 2012, pp. 53-57.

[3] Vo Thi Ngoc Chau; Nguyen Hua Phung, "A Knowledge-Driven Educational Decision Support System," Computing and Communication Technologies, Research, Innovation, and Vision for the Future (RIVF), 2012 IEEE RIVF International Conference, 2012, pp.1-6.

[4] D. L. Heisserer, and P. Parette, "Advising At-Risk Students in College and University Settings," College Student Journal, 36(1), 2002, pp. 69–84.

[5] G. L. Kramer, and M. McCauley, "High Tech and High Touch: Integrating Information Technology in the Advising Process," retrieved on December 10, 2010 from http://www.nacada.ksu.edu/Monographs/index.htm#m02.

[6] T. Feghali, I. Zbib, and S. Hallal, "A web-based decision support tool for academic advising," Educational Technology & Society, 14 (1), 2011, pp. 82–94.

[7] D. Yarbrough, "The engagement model for effective academic advising with undergraduate college students and student organizations," Journal of Humanistic Counseling, Education and Development, 41(1), 2002, pp. 61-68.

[8] T. M. Rao, S. Coleman, and C. Hollenbeck, "Advisor: An expert system for student advisement," Proceedings of the 15th annual conference on Computer Science, New York: ACM, 1987, pp. 32-35.

[9] Mansmann, S.; Scholl, M.H., "Decision Support System for Managing Educational Capacity Utilization," Education, IEEE Transactions, 50(2), 2007, pp.143-150.

[10] R.S. Pressman, Software Engineering: A Practitioner's Approach, 5thed, McGraw-Hill, 2009.

[11] H. Teo, L. Oh, C. Liu, and K. Wei, "An empirical study of the effects of interactivity on web user attitude," International Journal of Human-Computer Studies, 58, 2003, pp. 281-305.

Bangladeshi Style: a Way of Facial Artistic Stylization in Visual Effects

Md. Abul Kalam Azad
Department of
Computer Science and Engineering
ZH Sikder University of Science & Technology
Shariatpur, Dhaka, Bangladesh
akazad@zhsust.edu.bd

Ning Xie
Department of Computer Science
Tokyo Institute of Technology
Tokyo, Japan
xie@sg.cs.titech.ac.jp

Shamim Ahmad
Department of
Computer Science and Engineering
University of Rajshahi
Rajshahi, Bangladesh
shamim_cst@yahoo.com

***Abstract*—In computer graphics field, stroke-based painterly rendering (SBPR) is the program to convert photographs into paintings, such as oil painting, watercolor and sketches. Beyond specific techniques to render real photographs, major challenge is how to understand real photographs as human visual perception and cognition according to artists' painting motivation, which parts should be drawn and how to draw strokes. Before starting to draw with pigments, real artists commonly draw a sketch using simple lines to mark out final stroke layouts which navigate to control stroke drawing trajectories and poses. This can be treated as a blue print in their mind. In SBPR, it is also a powerful tool to guide strokes configuration. In this paper, we propose a image-based method to extract this blue print called guide-lines (GL). We generate GL by combining color gradient and morphological structure of real photographs. In the experiment, we apply our GL into Bangladeshi traditional mask automatic animation generation with webcam video. The results show that our method is sufficient and efficient to assist SBPR generation.**

***Index Terms*—Stroke-based painterly rendering, artistic stylization, GL extraction, visual effects.**

I. Introduction

The art of painting has long been a popular form of visual arts among people of different periods, places and cultures. Although the work of art may be represented in many different kind of types and art genres, the theories and practices behind these magnificent pictures are in common that reflects the human perception, psychological and emotional effects from the real world by the visual organs. Therefore, the study on the art theories and practices is nowadays of interest to not only the artists but also scientists and engineers. In the field of computer graphics, painterly rendering or artistic stylization is the method for rendering the real photo and video in a painterly style.

The avenue of digital era with its technological innovations, opens new perspective for the development of new art forms and technologies which are simply not possible by the conventional media. Nowadays, artistic stylization techniques are being employed worldwide in VFX (Visual effects) creation. However, this is an unexplored research area in Bangladesh. Our work aims to use visual effects in the context of Bangladesh.

The key motivation of our work is to apply artistic stylization in the context of Bangladesh following the existing state-of-theart technologies which would be an important contribution to the cultural digitalization of Bangladesh. But creating distinctive visual styles is a challenging task. Firstly, it requires to conform a coherent abstraction of particular expression and perception. Secondly, producing high quality artistic results usually have high computational cost.

Beyond the techniques for rendering the real photo into specific painterly styles, understanding the scene of the real photo is essential for automatic rendering. The approaches to understand real photos include the color distribution, the edge location, the region configuration, the texture information, and the spatial perception (foreground, background, and relations in perspective). The GL or painting composition is one of the powerful image understanding tool for stroke-based painterly rendering. It can navigate to control the trajectory and posture during drawing strokes by brushes. We focus our study to extract painting composition for human face stylization which is a simple, yet an important domain.

In this paper, we propose a framework to apply artistic stylization for human face that has many application areas such as entertainment and game industry. For example, the Bengali new year parade (Figure 1-a) is one of the most important traditional day and use of local masks on that day represents a cultural symbol in Bangladesh. So if we can digitalize the mask shown in (Figure 1-b), then during the festival, people can make a video call with the mask or can easily generate digital mask by using webcam only. Another application can be creating educational animation which would contribute to early childhood education in Bangladesh.

In our approach, we first locate the boundary of the facial components in a color image consisting of user's face using a deformable model fitting [1]. We then use an improved clustering approach to automatically generate efficient GL which guides a virtual brush agent [2] to stylized the user's face. The guide-lines can also be easily used for either the user interaction or the automatic method across the painting styles.

II. Related Work

Painting simulation and its application in VFX are now the state-of-the-art technology in the global community of com-

(a) (b)

Fig. 1. (a) Face mask on Bengali new year celebration is one of the most traditional cultural symbol in Bangladesh. (b) Magnified view of a face mask

.

puter graphics research but yet to explore in the perspective of Bangladesh. Past research studies have achieved many good results such as [2]–[4] in visual effects generation.

In the seminal work of Xie et al. [2], a virtual brush model was proposed that renders oriental brush strokes for painterly rendering of oriental painting style. We observe that this painting style had an influence on some famous Bangladeshi painting styles. Inspired from this work, our proposed techniques utilize the idea of painterly rendering as in [2] for facial stylization. In their system, the user inputs the contours of the complex shapes. After automatically estimating the brush trajectories, it renders them into brush strokes by mapping on them footprint textures scanned from real images. We employ this rendering approach in a novel way by improving the clustering method to best suit our purpose of human face stylization.

Lee et al. [3] presents a system for guiding free form drawing by dynamically providing a shadow image underlying the user strokes and by automatically blending relevant images from a large database to construct the shadows. The shadows are suggestive of object contours to guide the user as they draw. Grabli et al. [5] proposed an approach for clutter control in NPR (Non-Photo-realistic Rendering) line drawing where measures of view and drawing complexity drive the simplification or omission of lines. They offered a powerful and fine control on the final complexity, regularity, and style of the line drawing.

Further, Barla et al. [6] proposed an approach to the simplification of line drawings, in which a smaller set of lines is created to represent the geometry of the original lines. It maintains the morphological structure of the original drawing while allowing user-defined decisions about the appearance of lines. Our work is the natural extension of these methods of line drawing to the stroke-based drawing in the process of VFX generation.

Recently, Berger et al. [7] uses a data-driven approach to stylize and abstract human portrait which is quite similar to our idea of stylizing human face. However, they focus only on sketching of the human face in different level of abstraction. But our idea is to use brush strokes for painterly stylization of the human face.

III. The Framework

In this section, we provide the technical details of our framework for face stylization. Firstly, we detect the location of the facial components by [1] described in Section III-C. We then extract the Edge Tangent Flow (ETF) lines from the color image of the current frame of the webcam input. By clustering the similar curves of the ETF around the location of the facial components, we estimate the region of the the desired strokes. In the final stage, we render each region into strokes using a intelligent brush agent, modeled by Xie et al. [2] to get the desired stylization.

A. Oriental ink brush strokes rendering

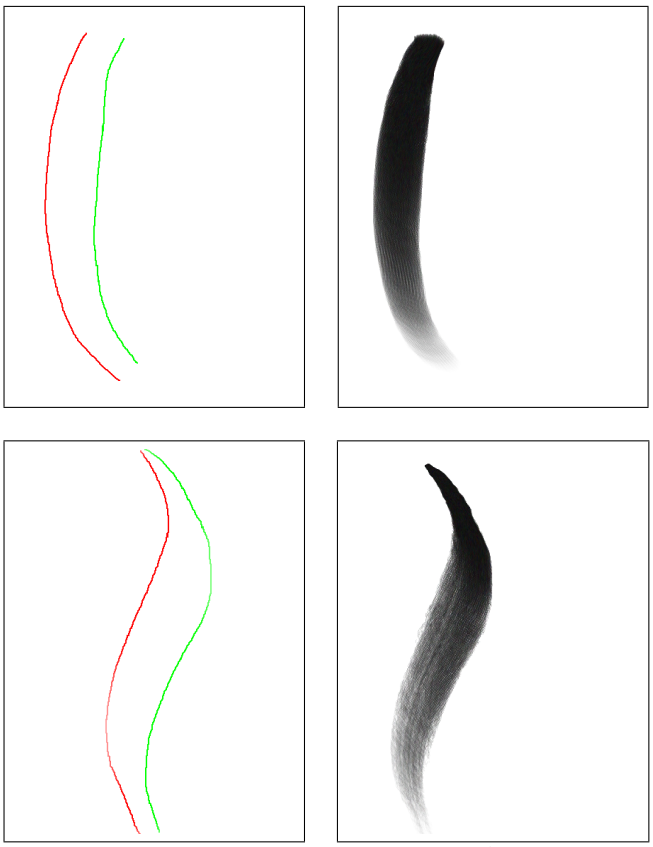
Oriental ink painting is a form of brush painting in which Chinese ink is applied to absorbent white papers. It is one of the oldest artistic brushwork and particularly very popular in Eastern Asian countries. Interestingly, we observe that the Chinese ink style had an influence on Zainul Abedin, who was a prominent Bangladeshi painter. For our purpose of face stylization, we use this style of rendering.

Given the painting composition, described in Section III-B, we aim to employ a brush model which automatically estimates the sizes of the brush footprint, their orientation and the brush trajectory to stylize human face. The approach operates as follows-

- *Optimal brush trajectory estimation* First complex shapes are automatically decomposed into elementary shapes that can be rendered with a single brush stroke. The decomposition is done by analyzing the distribution of branches of the medial axis of the shape. Second, by constraining the parameters related to the painting process, such as the variation of the footprint position, size, orientation, and angular velocity of the brush along the trajectory, the problem of optimal brush trajectory estimation is formulated as the minimization of an energy function that depends on these physical parameters.
- *Oriental ink painting style* In the second stage, the estimated trajectory is rendered into a brush stroke by mapping footprint textures scanned from real pictures [8].

The procedure synthesizes six basic stroke styles: dry ink, full-ink, hollow, first-half hollow, middle hollow, and both- ends strokes. Simulation of the pigment and water dispersion effects is done using Lattice Boltzmann Method [9] (LBM) to generate smooth and realistic strokes. The trajectory estimation stage provides a sequence of footprint locations, their sizes, and their orientations. The in-between footprints are generated using B-Spline interpolation [10]

The idea is that we assign the brush agent to draw a painting according to a map of desired strokes which we call GL as illustrated in Figure 2. In our work, we extract this map (GL) automatically using a improved clustering approach discussed in the next section.



(a) (b)

Fig. 2. Given the outlines of the brush strokes which we automatically extract for human face, plausible oriental ink brush strokes are rendered on these outlines [2]. (a) two basic stroke outline (b) corresponding rendered results

B. GL (Painting Composition) extraction

In the painting composition, the elements are a set of the boundaries of the candidate strokes, because we observe that the task of our applied brush agent [2] is to sequentially make actions to cover all regions one by one. We treat the painting composition as a segmentation problem for automatically collecting a set of the regions of the strokes. This requires segmenting the input image into a set of boundaries which will be used by our painting procedure for stylization. For this purpose, we take into account several useful information (such as ETF, region configuration) from the current frame image of a webcam video input. However, image segmentation is an open problem in image processing and computer vision.

1) Edge Tangent Flow (ETF): The vector field is a good tool to represent the color gradient. It can be used to predict the direction to draw the strokes according to the color gradient. We extract the Edge Tangent Flow (ETF) lines from the color image of the current frame from a webcam video input. Our purpose to use ETF is to guide the stroke contour extraction from the face image. To computer the ETF, we follow the same approaches as in ([2], [11])

We define the vector field $\boldsymbol{t}(\boldsymbol{X})$ from the input image $I(\boldsymbol{X})$, where $\boldsymbol{x}_i = (x_i, y_i)$ is an image pixel and $\boldsymbol{t}(\boldsymbol{x}_i)$ is a vector perpendicular to the image gradient $g(\boldsymbol{x_i}) = \nabla I(\boldsymbol{x_i})$. Thus, this vector filed is called an Edge Tangent Flow (ETF) since the vector field can be regarded as the tangent of a curve representing the local edge flow.

To generate the ETF, we use a kernel-based nonlinear bilateral filter [12] to iteratively optimize the vector field: $\boldsymbol{t}^{(i+1)} \rightarrow \boldsymbol{t}^{(i)}$ as

$$\boldsymbol{t}^{(i+1)}(\boldsymbol{x}) = \frac{1}{k} \sum_{\boldsymbol{y} \in \Omega(\boldsymbol{x})} \omega \phi(\boldsymbol{x}, \boldsymbol{y}) \boldsymbol{t}^{(i)}, \tag{1}$$

where k is the vector normalizing term. $\Omega(\boldsymbol{x})$ is the neighborhood of the pixel $\boldsymbol{x}$. ω is the weight value determined by the relationship with the neighbors as

$$\omega = \frac{1}{2} |\boldsymbol{t}^{(i)}(\boldsymbol{x}) \cdot \boldsymbol{t}^{(i)}(\boldsymbol{y})| (1 + \tanh(\hat{g}(\boldsymbol{y}) - \hat{g}(\boldsymbol{x}))), \tag{2}$$

where $\tanh(\cdot)$ is the target hyperbolic function, which is monotonically increasing and $\hat{g}(\boldsymbol{z})$ is the normalized gradient magnitude at $\boldsymbol{z}$.

2) Clustering lines: Given the generated stream lines after ETF construction and integration, We then cluster the similar curves (stream lines) from clusters of curves. First we sample at regular n_i points from each curve ψ_i. We say two neighbouring curves ψ_i and ψ_{i+1} are similar so that they can be put in the same cluster as

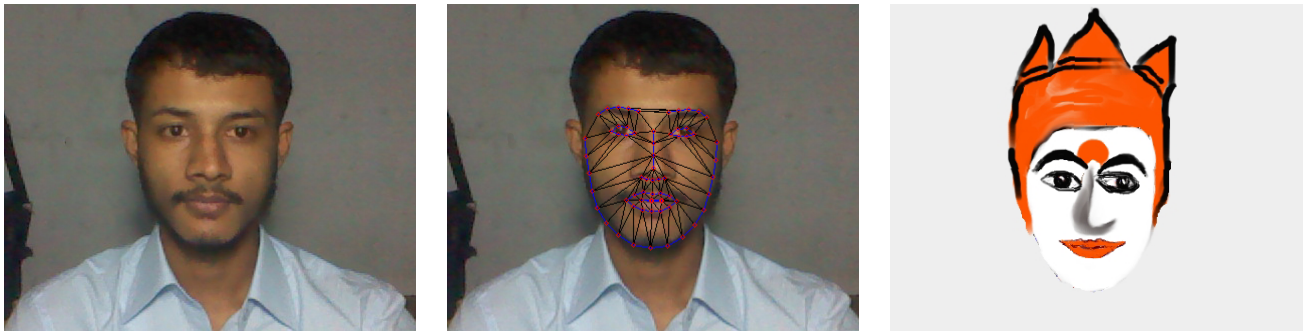
$$d_{SH}(\psi_i, \psi_{i+1}) < \varepsilon, \tag{3}$$

where d_{SH} is the *symmetric Hausdorff distance* defined between two sets of points by:

$$\begin{aligned} d_{SH}(\psi_i, \psi_{i+1}) &= \max(h(\psi_i, \psi_{i+1}), h(\psi_{i+1}, \psi_i)), \\ h(\psi_i, \psi_{i+1}) &= \max_{p \in \psi_i} (\min_{q \in \psi_{i+1}} \|p - q\|). \end{aligned} \tag{4}$$

In our experiments, we observed that the best clustering results are obtained when $\varepsilon = \sqrt{2}$.

Finally, we classify each pixel in the image to the cluster of its closest streamline. We then record the boundaries of the clusters as the boundaries of the candidate strokes in the painting composition.



(a) (b) (c)

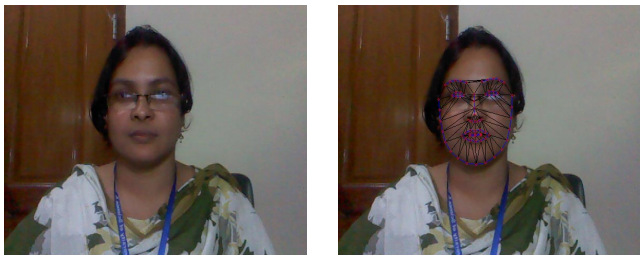
Fig. 3. (a) Input face image of a current frame, (b) extracted boundaries of the facial features, (c) simulated Bengali traditional style mask using our framework.

C. Facial key features extraction

In order to apply artistic stylization on facial image, we consider a current frame image consisting of the user's face in a video from web camera input. We also record face images consisting user's face under a set of different facial configurations. We extract the boundaries of the facial components such as mouth, nose, eyes, eyebrows and jaw from the current frame image by employing the algorithm in [1].

In our experiments, we use FaceTracker which is a C/C++ OpenCV implementation for real time generic non-rigid face alignment and tracking that is based on a modified version of the constrained local model described in [1]. It tracks the face from frame to frame. It detects the face using [13] framework and then as the face moves from frame to frame it just refines its guess using statistical deformable model fitting. But we aim to use it in artistic context.

FaceTracker allows us to track a face by locating a face mesh (consisting of 68 2D landmark points) to recognize the edge of lips, jaw, nose, eyes and eyebrows. By observation, we then label these landmark points into different groups. For example, landmark points from 0 to 16 belong to chin boundary, 22 to 26 belong to right eyebrow boundary, from 42 to 47 belong to the right eye boundary, and from 36 to 41 belong to the left eye boundary.



(a) (b)

Fig. 4. Boundary extraction of facial components (right) such as eyes, mouth and chin of the input frame (left) to generate guide-lines for different facial configurations (male/female). Later, we use the GL results to guide the face stylization.

These extracted boundaries will be required by the rendering algorithm which we discussed in Section III-B for face stylization. Figure: 4 shows some of our extracted key feature's boundary results for different facial configurations.

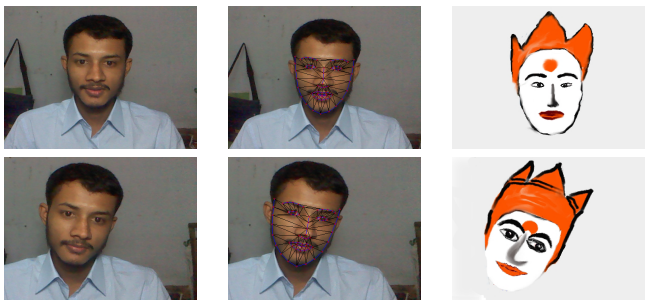


Fig. 5. Results of simulated traditional style mask using our framework for other facial configurations. More results can be found on: https://sites.google.com/site/seanxiening/projectbdstyle

IV. Conclusion and Future Work

We have presented a framework that extracts efficient GL or painting composition from an input image. The framework is used later to stylize human face. We believe that our extracted painting composition could be utilized across different painting styles as well. Currently, we consider frame by frame off-line VFX generation from a webcam video input and focus on a specific domain, human face stylization. We have rendered a traditional Benglai mask avatar (Figure 3 and 5) which is based on physical simulation. We also consider force to model the face motion. In our experiments, all the example avatar mask rendered are generated within 2∼10 minutes including the rendering time.

In the future work, we plan to extend our idea to add human face emotion detection and then control the emotion stylization in the mask avatar on real-time.

References

[1] J. M. Saragih, S. Lucey, and J. F. Cohn, "Deformable model fitting by regularized landmark mean-shift," *Int. J. Comput. Vision*, vol. 91, no. 2, pp. 200–215, Jan. 2011. [Online]. Available: http://dx.doi.org/10.1007/s11263-010-0380-4

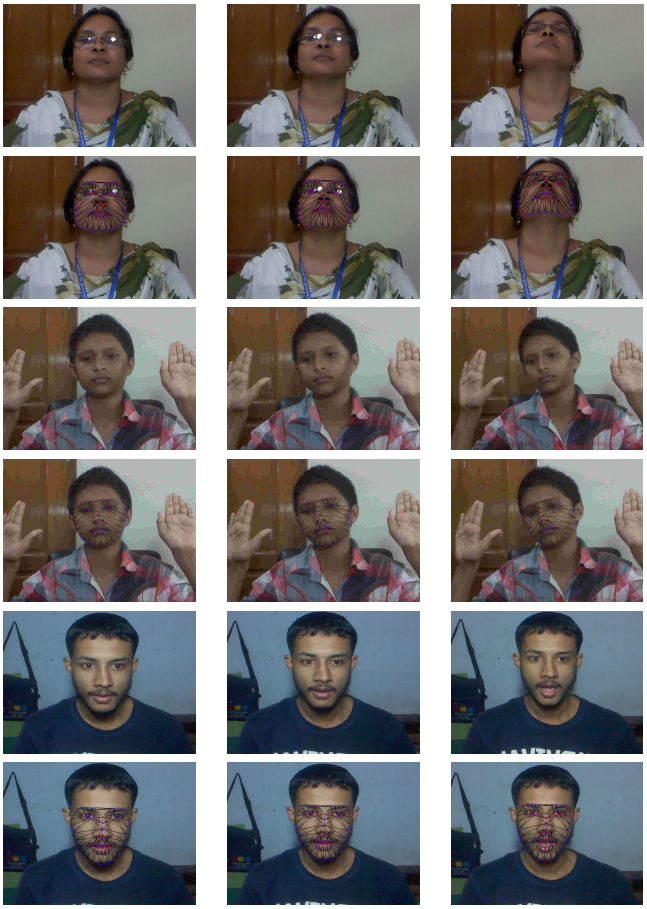


Fig. 6. Boundary extraction for different facial expression on different subjects (i.e. female, kid, male). First row of each subject contains the input frame image of different poses and the next row shows the corresponding extracted facial components. We would like to consider VFX generation in real-time for these different facial expression in future work.

[2] N. Xie, H. Laga, S. Saito, and M. Nakajima, "Contour-driven sumi-e rendering of real photos," *Computers & Graphics*, vol. 35, no. 1, pp. 122–134, 2011.

[3] Y. J. Lee, C. L. Zitnick, and M. F. Cohen, "Shadowdraw: real-time user guidance for freehand drawing," *ACM Trans. Graph.*, vol. 30, no. 4, pp. 27:1–27:10, Aug. 2011. [Online]. Available: http://doi.acm.org/10.1145/2010324.1964922

[4] D. DeCarlo and A. Santella, "Stylization and abstraction of photographs," in *Proceedings of the 29th annual conference on Computer graphics and interactive techniques*, ser. SIGGRAPH '02. New York, NY, USA: ACM, 2002, pp. 769–776. [Online]. Available: http://doi.acm.org/10.1145/566570.566650

[5] S. Grabli, F. Durand, and F. Sillion, "Density measure for line-drawing simplification," in *Proceedings of Pacific Graphics*, 2004. [Online]. Available: http://maverick.inria.fr/Publications/2004/GDS04

[6] P. Barla, J. Thollot, and F. Sillion, "Geometric clustering for line drawing simplification," in *Proceedings of the Eurographics Symposium on Rendering*, 2005. [Online]. Available: http://maverick.inria.fr/Publications/2005/BTS05a

[7] I. Berger, A. Shamir, M. Mahler, E. Carter, and J. Hodgins, "Style and abstraction in portrait sketching," *ACM Trans. Graph.*, vol. 32, no. 4, pp. 55:1–55:12, Jul. 2013. [Online]. Available: http://doi.acm.org/10.1145/2461912.2461964

[8] M. D. S. S. N. M. Okabe, Yuta, "An oriental rendering of lines by using discrete hmm," in *IEICE J90-D(1)*. The Society for Imaging Science and Technology, 2007, pp. 106–114.

[9] N. S.-H. Chu and C.-L. Tai, "Moxi: real-time ink dispersion in absorbent paper," in *SIGGRAPH '05: ACM SIGGRAPH 2005 Sketches*. New York, NY, USA: ACM, 2005, p. 62.

[10] H. S. Seah, Z. Wu, F. Tian, X. Xiao, and B. Xie, "Artistic brushstroke representation and animation with disk b-spline curve," in *ACE '05: Proceedings of the 2005 ACM SIGCHI International Conference on Advances in computer entertainment technology*. New York, NY, USA: ACM, 2005, pp. 88–93.

[11] H. Kang, S. Lee, and C. K. Chui, "Coherent line drawing," in *NPAR '07: Proceedings of the 5th international symposium on Non-photorealistic animation and rendering*. New York, NY, USA: ACM, 2007, pp. 43–50.

[12] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Proceedings of the Sixth International Conference on Computer Vision*, ser. ICCV '98. Washington, DC, USA: IEEE Computer Society, 1998, pp. 839–. [Online]. Available: http://dl.acm.org/citation.cfm?id=938978.939190

[13] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," 2001, pp. 511–518.

Security Risk Modelling Using SecureUML

Mohammad Jabed Morshed Chowdhury
Department of Computer Science and Engineering, Daffodil International University
Dhaka, 1209, Bangladesh
jabedmorshed@gmail.com

***Abstract*— Several security modelling languages (e.g., Misuse case, Secure Tropos) help dealing with security risk management at the system requirements stage. But no design level modelling language has been explored to model security risk. In this paper, we are focusing on SecureUML which is a design level modelling language to represent security risk. More specifically we investigate how SecureUML supports information systems security risks management (ISSRM). The outcome of this work is an alignment table between SecureUML language constructs to the constructs of the ISSRM domain model. We ground our analysis on the number of illustrative examples. We hope that our results will help developers to understand how they can consider security risks at the system design stage. It also indentifies the shortcomings of SecureUML to model security risk and provides recommendations for improvement.**

***Keywords*— Security, Security risk modelling, SecureUML.**

I. Introduction

Nowadays, Information Systems (IS) have large impact on our personal, social and economical life. Facebook, Twitter and e-Commerce sites are common phenomena in our daily life. Introduction of information system has made our life easy, convenient and more relaxing. But like many other opportunities, information system also has its drawbacks. Hacking of websites and stealing credit card and personal information are not rare in Internet world. To undermine all kind of potential attacks IS needs to be more secured. The significance of security technologies is now widely accepted. As a result security facets play a vital role in any modern information system.

Though security becomes a buzz word for information technology world but it is often seen as ad-hoc basis during the later part of the development cycle [1, 2]. This approach (ad-hoc) often leads to errors and vulnerabilities that provide a potential for exploitation. Statistics from CERT/CC shows that the number of security incidents has increased quite significantly from 2000-2007 (Fig. 1).

A large number of software vulnerabilities have indicated that these systems are repeatedly lacking the appropriate measures from design and implementation perspective. Any minor flaw at the design stage can contributes significantly to application vulnerabilities, which can be exploited by the attackers. A study [3] shows that design stage problems were accounted for about 50% of the security flaws during the Microsoft's "security push" in 2002. Viega and McGraw [4] have mentioned design stage vulnerabilities as the second major cause of security risk in software in their book.

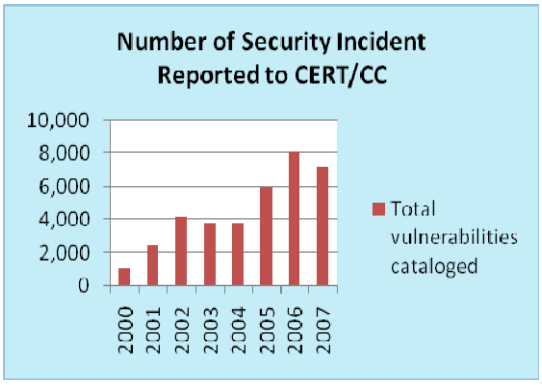


Fig. 1. Number of security incident reported to CERT/CC

By following an improved and well structured process at design stage, with security in mind can reduce these vulnerabilities quite significantly. Several security modelling languages (e.g., Misuse case, Secure Tropos) have been investigated for dealing with security risk management at the requirements stage. Mal-activity diagram has been examined as the post requirement stage modelling language [8]. But no design stage language modelling language has been investigated yet to model security risk. Thus this work is motivated to help developers to understand how they can model security risks at the system design stage. More specifically, SecureUML is used to model security risks. This work is aimed to show how SecureUML can be used to model security risk. It also finds out its limitations and proposes the solutions for improvements.

The structure of this paper is as follows: Section 2 introduces the ISSRM domain model, which is the basis for analysing SecureUML. Section 3 illustrates how SecureUML could be applied for security risk management. Section 4 describes how SecureUML constructs are aligned to the concepts of the ISSRM domain model. Section 5 presents the lessons learnt.

II. Background

Information System Security Risk Management (ISSRM) [5, 6] is a systematic approach, which addresses the security related issues in an IS domain. The model is defined after a survey of risk management and security related standards,

security risk management methods and software engineering frameworks [6]. The domain model (see Fig. 2) supports the alignment of security modelling languages. It improves the IS security and security modelling languages as it conforms to the security risk management of organizations. The model describes three different conceptual categories:

Asset-related concepts describe the organization's assets grouped as business asset and IT asset. It also defines the security criterion as a constraint of a business asset expressed as integrity, confidentiality and availability.

Risk-related concepts define risk, potential harm to business, it is composed of a threat that contains one or more vulnerabilities, if executed successfully, harms the system assets which has negative consequences on assets defined as an impact. They negate the security criterion imposed by the business asset. An event is an abstraction aggregated as a threat and vulnerability where vulnerability is a weakness in a system that can be exploited by threat agent. A threat is a way to inflict an attack. It harms IS and business asset carried out by a threat agent and an attack method to target IS assets. Threat Agent is an attacker that initiates a threat to harm the IS asset. Attack Method is a mean through which a threat agent executes a threat.

Risk treatment related concepts define a risk treatment decision to avoid, reduce, retain, or transfer the potential risks. It is refined by the security requirement. A control implements the security requirement. The ISSRM process [5] is a 6-step process, based on existing risk analysis methodologies and standards. It starts with context and asset identification of the organization, proceeding to determine the security objectives for identified assets. Next, risk analysis and assessment to examine and estimate potential risks and its impacts. In next step, risk treatment decisions are taken to identify the security requirements. Finally, security control is implemented as security requirement. The process is iterative which may identify new risks and security controls.

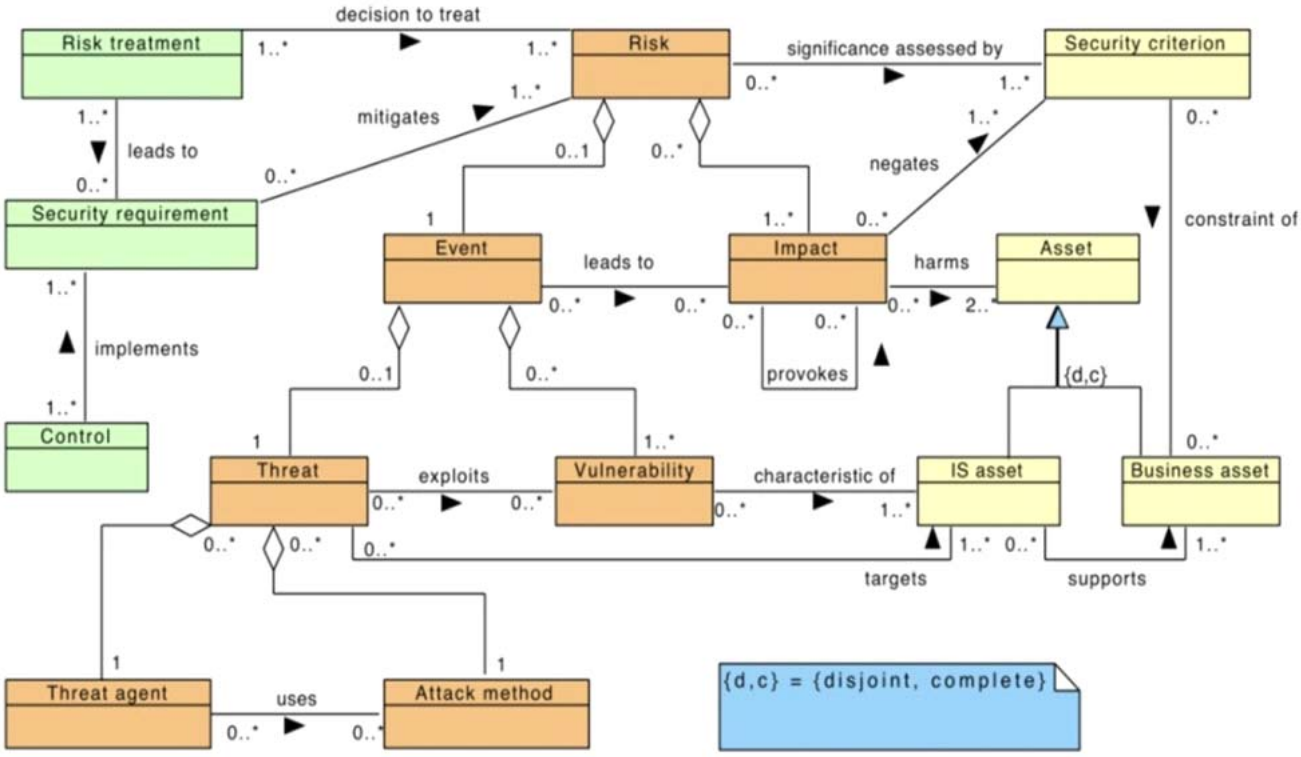


Fig. 2. The ISSRM Domain Model

III. SecureUML

SecureUML [7] is based on role-based access control with additional support for specifying authorization constraints. It embeds role-based access control policies in UML class diagrams using a UML profile. It defines a vocabulary for annotating UML based models with information relevant to access control. SecureUML defines a vocabulary for expressing different aspects of access control, like role, role permission and user-role assignment.

Authorization constraint defines the precondition for granting access to any operation. Such constraints are expressed using Object Constraint Language (OCL) [1]. Authorization constraint gives SecureUML the flexibility to define and verify the access decision on dynamically changing data. Like UMLSec, SecureUML is also used at the design

[1] http://www.cs.ucl.ac.uk/staff/ucacwxe/lectures/3C05-04-05/OCL-Essay.pdf

stage of the software development process. We will thoroughly investigate different concepts of SecureUML and will also try to find the concept alignment with ISSRM domain model in section IV.

We will use SeureUML on online banking system described in [8]. SecureUML uses its main concepts (e.g., user, role, component, and permission) to design software application. Figure 4 shows a scenario of money transfer. SecureUML is used by following steps,

1.) **Identify Users.** In this system Alice is the client of this online bank. So, Alice can be seen as the user (a class with a stereotype `<<user>>`) of this system.

2.) **Identify application roles.** Online banking system defines a role (a class with a stereotype `<<role>>`) account holder for the client to get the service from the system.

3.) **Map users into roles.** User Alice is mapped with the role account holder in the system.

4.) **Identify resources**. Bank account (a class with a stereotype `<<ModelElement >>`) is treated as the resources for this example.

5.) **Identify actions.** Stereotype permission (an association class with stereotype `<<permission>>`) is used to define the action for the system. It defines different actions by using attributes action type. In this example, read and update can be as action.

6.) **Identify authorization constraints**: authorization constraints must be satisfied in order for an authorization to apply. It can be seen as the precondition. In this example, transaction can be done in business hours only. It is ensured by the constraint `BusinessHoursOnly`.

7.) **Account for cardinality.** SecureUML also shows cardinality. Here it shows one-to-many relationship between Account holder and Bank account.

3.1 Meta-Model Of SecureUML

The SecureUML meta-model (Figure 3) is represented by the Role Based Access Control (RBAC) concepts. It introduces three new concepts *User, Role* and *Permission*. Due to the design decision every UML model element is represented as a protected resource and defined as `ModelElement`. ResourceSet is used to which represent a user defined set of model elements. Permission is a relation object connecting a role to a `ModelElement` or a ResourceSet. The semantics of permission is defined by the ActionType elements used to classify the permission (see the association ActionTypeAssignment). Every Action-Type represents a class of security relevant operations on a particular type of protected resource. Action types give the developer a vocabulary to express permissions at a level close to the domain vocabulary.

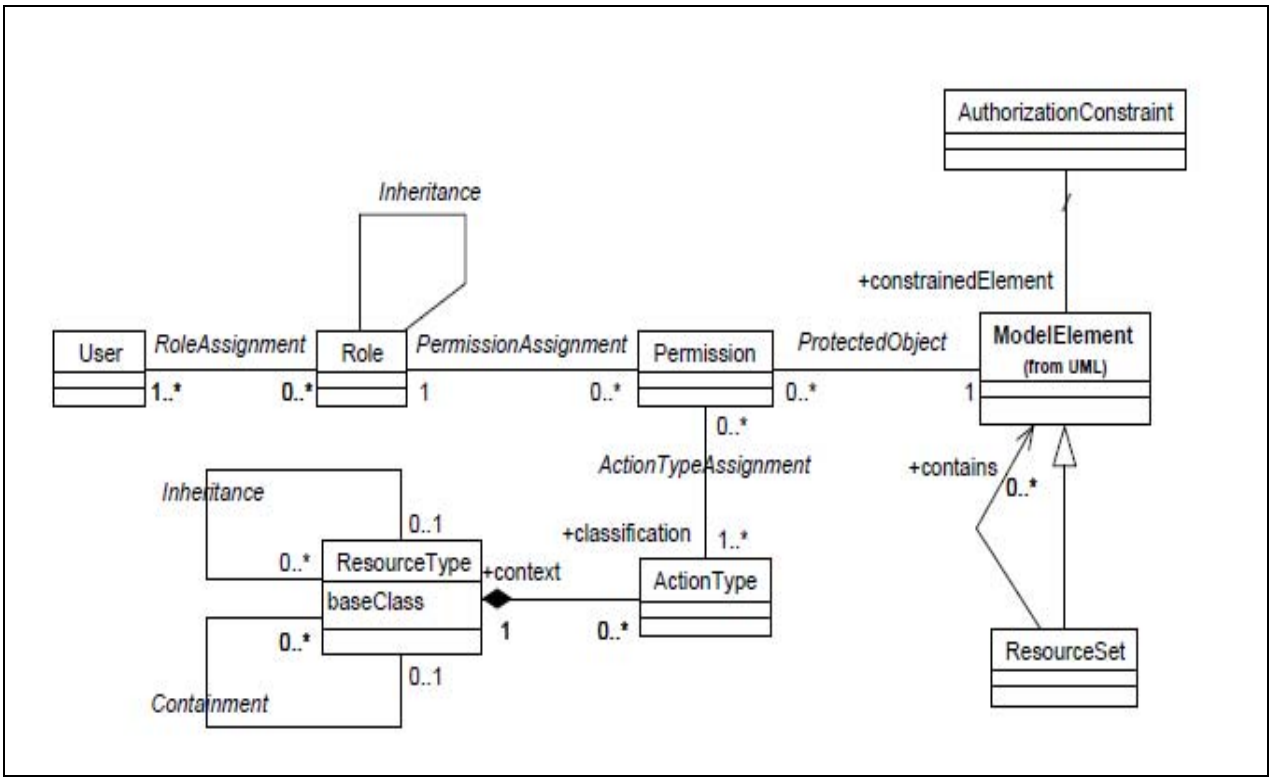


Fig. 3. SecureUML Meta-model

The set of action types available in the language can be freely defined using ResourceType elements. A ResourceType defines all action types available for a particular meta-model type. The set of resource types and their action types help to define the resource type model for the platform. An `AuthorizationConstraint` is a part of the access control policy of an application. It expresses a precondition imposed on every call to an operation of a particular resource, which usually depends on the dynamic state of the resource, the current call, or the environment.

IV. Alignment of SecureUML with ISSRM

Here we will follow ISSRM process and use SecureUML constructs to analyse the online banking scenario.

(a) Context and Asset Identification. In Fig. 5 we illustrate the context of online bank. We assume there are two roles: Client and Bank (represented as class with a stereotype `<<Role>>`) used for bank account control management. The client can access BankAccount (class with a stereotype `<<ModelElement>>`) and can transfer money by using its operations (e.g., deposit (), withdraw ()). The attribute (e.g., balance) of the Bank account shows the current balance of the account. The authorization on *BankAccount* for Client is assigned by *ClientPermission* (association class with a stereotype `<<Permission>>`). The client can read and update (ActionType) the `BankAccount` information. Like the client role, Bank officer role also authorized to the account by Permission (*BankOfficerPermission*). Bank officer can perform read, create and delete (ActionType) operation on *BankAccount*. One client can have more than one account and one account can be managed by more than one Bank officer. Bank officer also can manage many accounts.

Here, *BankAccount* class holds the client's banking information which can be seen as ISSRM asset. The client is concern about the manipulation of the balance information which represents the business asset. Integrity of the balance information can be defined as security criterion.

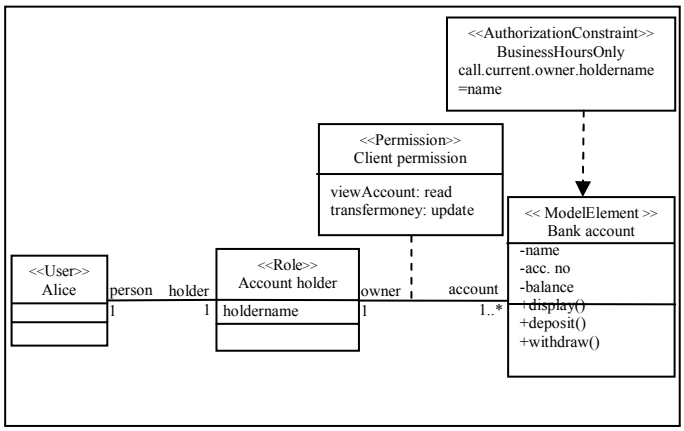


Fig. 4. SecureUML model for a money transfer

(b) **Security objective determination**. Security objectives are often defined in terms of confidentiality, integrity and availability. Here we are concerned about the balance information. We focus on the integrity of the balance information because we care that it would be manipulated in non-intended way.

(c) **Risk analysis and assessment**. A malicious actor (e.g., hacker) can steal the login information by using email system and malware. We are assuming that the hacker has already received the login information (e.g. as described in Fig. 6). Where the hacker will login to the system with the stolen login information and will get the same access level as Client. Now the hacker will have all kinds of operation authorization (*Permission*) to the Bank account (*ModelElement*) as Client. He can now transfer (*action type*) money to his own account. ISSRM helps to identify what access is allowed or not for different roles within the SecureUML model.

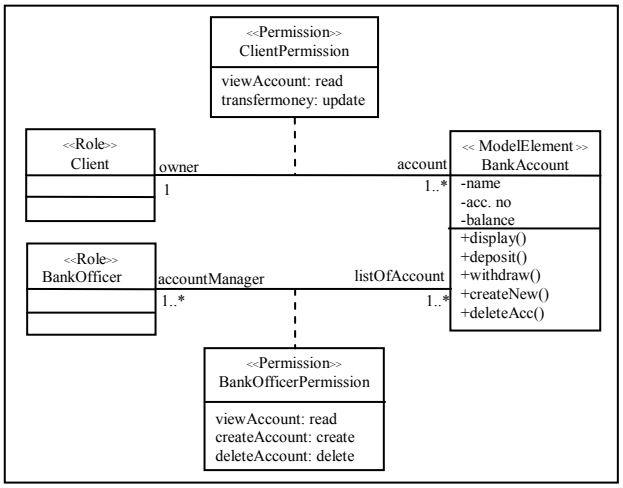


Fig. 5. Asset modeling by SecureUML

(d) **Risk treatment.** In this step we choose one out of four risk treatment procedures. Here we are choosing risk reduction (taking actions to lessen the probability and negative consequences of the risk) rather than risk avoidance (decision is taken for not to get involved in any risk situation). Here we are not taking any decision to avoid the situation but trying to improve the security measure to lessen its probability and effect.

(e) **Security requirements definition.** By setting *AuthorizationConstraint* we can put our security requirement. Here constraint *OneTimePassword* will ensure that even if the login information is compromised the malicious actor (Hacker) cannot transfer (update action type) money to other account.

Every client is given a printed one time code book. They use this code as the confirmation of any create, update or delete operation on Bank account. For example, this code is used to confirm the money transfer. The main feature of this code is that it can be used only once and after that it becomes invalid. The banking system checks this code with the code book for that specific client. So, even if the hacker steal the login information he can only see (read action type) the account information. If he tries to transfer money to his account he will be prompted for the one time code as confirmation. As he does not have access to this one time code book he could not transfer this money to his account.

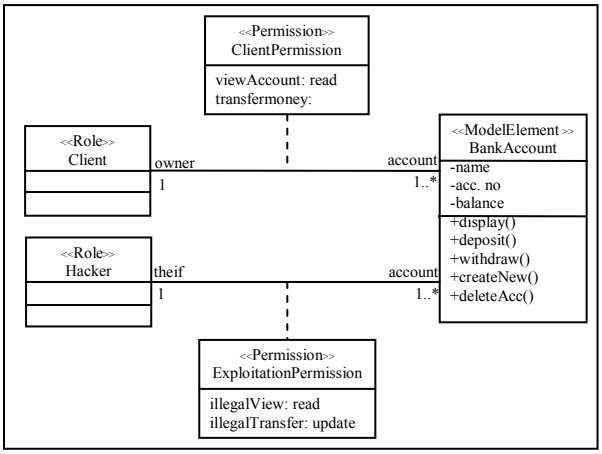


Fig. 6. Threat modelling by SecureUML

(f) **Control selection and implementation.** From this model we can transform to implementation [9, 10]. The overall model defines rather a control mechanism and links it to the secured assets than separate concepts of ISSRM.

4.2 SecureUML and ISSRM Reference Model

(a) **Asset-related Concepts.** ISSRM asset is anything that is valuable and vital for achieving the origination's goal. Business asset is defined as the information, process, plan and skill essential to the business. In SecureUML, *ModelElement* (*BankAccount*) correspond to protected resources. *ModelElement* also includes *ResourceSet* which represents a user defined set of model element. Protected resources represent the module that is valuable for the organization to provide service. Thus *ModelElement* can be mapped with asset. The attribute (*balanec*) of the *ModelELement* represents the information that is critical for the client as well as for banking system.

A *Role* is a job or function within an organization. It combines all privileges needed to fulfil the respective job or function. It represents the way how the user uses the system. *Permission* represents the authorization to execute an operation on one or more protected objects or resources. The semantics of *permission* is defined by the *ActionType* which represents a set of security relevant operation on protected resource. It helps the user to access the protected vital information. The set of action types for the meta-model is defined by the *ResourceType*. We can categorize Role, *Permission, ActionType* and *ResourceType* as the element which helps to protect Bank account (*ModelElement*). The operations (e.g., withdraw, deposit) of the Bank account is the mean which helps to work on the balance information. In this perspective, these constructs can be aligned with IS asset. SecureUML is focused on security and resources (*Bank account*) can be seen as protected resource. But is does not contain any constructs for explicit definition of the ISSRM security criterion.

(b) **Risk-related Concepts**. It discusses about the risk and its consequences. In principal SecureUML does not have the support to represent security risk. "at the methodological level SecureUML only focus on the solution domain." [9].

However, we have used SecureUML to define the threat model (Figure 7) of the running example to show the negative/harmful effect. Here *hacker* (*role*) which takes the unlawful benefit by transferring money can be aligned with threat agent. The means used by the threat agent to execute threat is attack method. The attack method (e.g., *illegalView: read, illegalTransfer: update*) can be aligned with the attribute of an association class *Permission*. Hacker uses these permission to exploit the vulnerability and targets the IS assets. Vulnerability of the IS asset is defined implicitly (no security confirmation mechanism before operation is performed). Exploitation of this vulnerability will result into an impact (Loss of integrity of the money transfer) which will cause harm to the asset. We can implicitly identify the vulnerability and impact but SecureUML does not have any concrete construct to represent them. SecureUML also does not have any construct to define risk and event.

(c) **Risk Treatment-related Concepts.** It deals with the risk treatment decision and implementation of control in order to mitigate risk (Figure 7). Here we have selected risk reduction as a risk treatment decision for our example. ISSRM security requirement is represented by *AuthorizationConstraint* (*OneTimePassword*). It is attached with the *ModelElement* via *constrainedElement.*

Table 1 shows the concept alignment of SecureUML and ISSRM concepts. Under the SecureUML column we write down the graphical constructs of SeureUML that relates to the ISSRM concepts and in the next column we have shown the example.

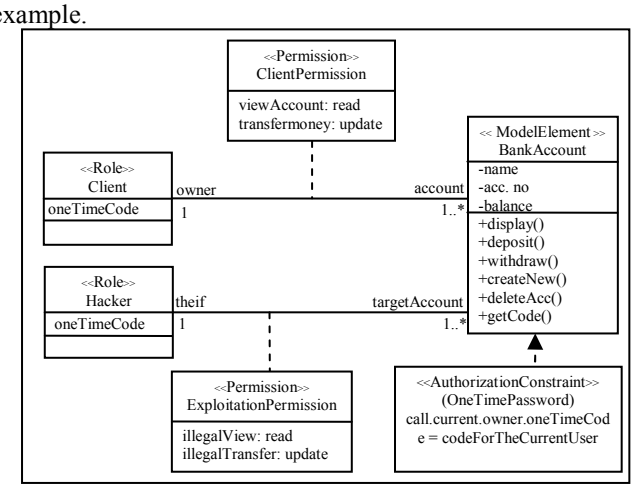


Fig. 7. Modelling of security requirements

TABLE I. Concept Alignment Between SecureUML and ISSRM

ISSRM domain model		SecureUML	
		SecureUML	Example
Asset related concept	Asset	Class with a stereotype <<ModelElement>>	Bank account
	Business asset	Attributes of the class stereotype <<ModelElement>>	balance, name, acc. no
	IS asset	Class with a stereotype <<Role>>, Association class with a stereotype <<Permission>>, Operations of the class stereotype <<ModelElement>>	Client, Bank Officer Client permission, Bank officer permission display(), deposit(), withdraw(),viewAccount: read, createAccount: create, deleteAccount: delete, transfermoney: update
	Security criterion	--	Integrity of the balance
Risk related concept	Risk	--	--
	Impact	--	Loss of integrity of the money transfer
	Event	--	--
	Vulnerability	--	No security confirmation mechanism before operation is performed
	Threat	Class with a stereotype <<Role>>, Association class with a stereotype <<Permission>>	Transfer money by malicious actor
	Threat agent	Class with a stereotype <<Role>>	Hacker
	Attack method	Attributes of an association class with a stereotype <<Permission>>	illegalView: read illegalTransfer: update
Risk treatment related concept	Risk treatment	--	Risk reduction
	Security requirement	Authorization constraint, constrainedElement	OneTimePassword
	Control	The whole model as control	Bank account control management

V. DISCUSSION AND CONCLUSION

In this paper we have shown the concept alignment between ISSRM reference model and SecureUML. We have analysed the concepts by using a running example. The final outcome of this paper is an alignment table (Table 1). This work primarily shows that how we can use SecureUML to model the security risk at design stage. Our analysis also shows that SecureUML has several limitations to define the security risk management concepts. Here we will suggest few improvements for SecureUML from security risk management prospective.

- Several concepts (e.g., risk, event) do not have any construct to represent security risk. One solution could be to introduce new stereotypes and another one could be to define tagged values within existing stereotypes. Same like as we have proposed for Mal-activity diagram in another work [11].
- Sometimes same constructs (e.g., Role, Permission) is represented for different concepts (e.g., IS asset and attack method). Specific tagged value could be used to differentiate the same constructs (Stereotype).

This work has identified some limitations of SecureUML to model security risk. The major future work remains implement of the identified extension to the modelling languages. Before this implementation could happen, validation should also be done by security experts. Our analysis also opens the way for the interoperability between different modelling languages that are analysed using the same conceptual background, thus, potentially leading to the transformation between different modelling approaches

ACKNOWLEDGMENT

I would like to acknowledge Dr. Raimundas Matulevicius (Estonia) for his support, guidance and encouragement to carry out this work. I am also grateful to Guttorm Sindre (Norway) and Péter Kárpáti (Norway).

REFERENCES

[1] Mayer N., "Model-Based Management of Information System Security Risk". Doctoral Thesis in Computer Science Namur, Belgium, 2009

[2] Sindre G., "Mal-Activity Diagrams for Capturing Attacks on Business Processes". In proceedings of the Working Conference on Requirements Engineering: Foundation for Software Quality, 2007.

[3] Hoglund G. and McGraw G., "Exploiting Software: How to Break Code". Book published from Addison-Wesley, 2004. ISBN: 0-201-78695-8.

[4] Viega J. and McGraw G., "Building Secure Software: How to Avoid Security Problems in the Right Way". Book published from Addison Wesley, 2002. ISBN: 0201-72152-X.

[5] Mayer N.: Model Based Management of Information System Security Risk. Doctoral Thesis in Computer Science, University of Namur, Belgium (2009)

[6] Mead N. R., Hough E. D. and Stehney T. R. II.: Security Quality Requirements Engineering (SQUARE) Methodology. Technical Report CMU/SEI-2005-TR-009, Carnegie Mellon University, Software Engineering Institute, Pittsburgh, Pennsylvania (2005)

[7] Lodderstedt T, Basin D. and Doser J., "SecureUML: a UMLbased modeling language for model-driven security". In Proceedings of the 5th International Conference on the Unified Modeling Language, 2002.

[8] Chowdhury M. J. M., Matulevicius R., Sindre G., Kárpáti P.: Aligning Mal-activity Diagrams and Security Risk Management for Security Requirements Definitions. REFSQ 2012: 132-139

[9] Matulevicius R., Lakk H. and Lepmets M., "An approach to assess and compare quality of security models". ComSIS, Volume 8 No 2, Special Issue, May 2011.

[10] Matulevicius R. and Dumas M., "A Comparison of SecureUML and UMLsec for Role-based Access Control". In proceedings of the 9th International Baltic Conference, University of Latvia Press, 2010.

[11] Chowdhury M. J. M. "Towards Security Risk-oriented Mal Activity Diagram". International Journal of Computer Applications 56(10):47-52, October 2012. Published by Foundation of Computer Science, New York, USA.

Machine Understandable Information Representation of Geographic Related Data to the Administrative Structure of Bangladesh

Md. Hasan Hafizur Rahman
Dept. of Computer Science & Engineering
University of Chittagong
Chittagong - 4331, Bangladesh
Email: hasancse03@gmail.com

Shima Chakraborty
Dept. of Computer Science & Engineering
University of Chittagong
Chittagong-4331, Bangladesh
Email: shima_csecu@yahoo.com

Md. Hanif Seddiqui
Dept. of Computer Science & Engineering
Bangabandhu Sheikh Mujibur Rahman Science and Technology University, Gopalganj-8100
and University of Chittagong, Chittagong-4331
Email: hanif@cu.ac.bd

***Abstract*—The heterogeneous data on the web related to the geographic information of a country is an increasingly important field of data sharing and integrating with diverse sources for retrieving suitable information using search engines. These heterogeneous geographic data is mostly available in unstructured or semi-structured format. Converting and integrating these data into a machine understandable representation is a challenging task and is getting researchers' attention at a rapid pace. In this regard, we described geographic data for administrative structure of Bangladesh which integrated all resources and instances by comprising of their concepts and relations, coined a repository name *Geo-Bangladesh* by rich semantic Resource Description Framework (RDF). Then we utilized our *Geo-Bangladesh* with one more knowledge repository *Bangladeshi-Citizen* to achieve the semantic interoperability in our application. Furthermore, we performed SPARQL query operations on our proposed semantic knowledge repository to retrieve and inference specific information that shows its usability and effectiveness through the adequate quantitative results, both in term of concepts and geographical entities of the administrative structure of Bangladesh.**

I. Introduction

The procreation of heterogeneous data sources of geographic information along with the expeditiously growing World Wide Web (WWW) in an unstructured format with human understandable representation are insufficient for even smarter method in finding specific information on the web. These heterogeneous data repositories in the form of document having unstructured text related to particular domains are easily readable by humans, however it is hardly understandable to machine. To develop unstructured web contents we use Hypertext Markup Language (HTML)[1], Dynamic HTML (DHTML)[2] where a sequence of characters declaring by predefined tags indicating the appearance of contents in the document instead of their machine understandable semantics. These tags are used to affirm the style rather than the interpretation of data in the documents. The following tag representation display *Hafizur Rahman* in a paragraph of a document without explicit meaning of $Hafizur\ Rahman$ instead of a sequence of characters.

e.g, $< p > Hafizur\ Rahman < /p >$

[1]http://www.w3.org/html
[2]http://en.wikipedia.org/wiki/Dynamic_HTML

Therefore, Extensible Markup Language (XML)[3] is a semi-structured data representation technology having user defined tags. By reason of the inadequacy of uniqueness in user defined tags to represent information in a document is scarcely understandable to machines. In spite of these expeditiously growing data using these technologies, current WWW also contains a lot of database driven contents usually called as dynamic web where we design database schema first and then we add the contents. Although these contents are structured and rely on their own relational database systems, they are usually lack of machine understandable semantics. As a result of the deficiency of semantics, the traditional search engines do not attain specific information on the web. Consider a user query related to the place $Dhaka$, *capital city of Bangladesh* traditional search engine retrieves $Dhaka$ related data without interpretation $Dhaka$ is a place or not by searching different scattered and unstructured sources on the web, then feedback relevant and irrelevant documents to search keyword as well.

Moreover, different software of Geographic Information System (GIS) uses incompatible system designs, data models, and database storage structures those are composed into a proprietary format [1], [2], [3]. The integration efforts of geographical information is a formidable task mainly rely on these proprietary solutions. Furthermore, geographical data semantics those deal with representations and reasoning on the meaning of geospatial data are critical for the development of interoperable geographical data and software [4], [5], geographical information retrieval [6], and automated spatial reasoning [7]. However, it is extremely difficult to capture and maintain semantic knowledge of geographical data due to the complexities of geographical categories [8], geospatial languages [9], and heterogeneous representation of spatial data. A series of specifications such as Geographic Markup Language (GML) [10] and data access protocols such as Web Feature Service (WFS) [11], and Web Map Service(WMS) [12] are able to access heterogeneous contents stored in the web. These standards do not have enough constructs to express data semantics on the web. On the other hand, ontology is used to define a common vocabulary that minimizes the semantic problems in interoperability, metadata modeling, and communicating

[3]http://www.w3.org/XML/

meaning of data across domains, and data integration [13], [14], [6]. It is employed as a method for identifying categories, concepts, relations, and rules that prescribe theories of the geographical domain [15]. However, a lot of efforts has been made to develop the next generation web, *Semantic web*, is a term coined by Berners-Lee et al. [16], define the objective of data on the web to achieve the semantic interoperability among metadata associated with the web information. In spite of the growth of semantic web technology that addresses the issue of data integration and machine understandability of huge data repositories, geographic information of a country is a motivated complication problem with vast application areas such as seeking related distinguishing information from interoperable web of data. In this regard, the main center of attraction of our research is to develop a generic machine understandable database $Geo-Bangladesh$ from the existing available data using rich semantic web technologies to achieve resource sharing and interoperability issues in a faster and efficient way. We performed a couple of experiments to retrieve and inference specific information using a SPARQL that shows effectiveness of our proposed knowledge repository for geographic information of Bangladesh. The interoperability issue is cross-examined with knowledge repository $Bangladeshi-Citizen$ related to the citizen information of Bangladesh.

The rest of the paper is organized as follows. Section II introduces general terminologies to comprehend consequent contents of this paper. Our present administrative structure of Bangladesh is articulated in Section III, while Section IV focuses on the details data preparation procedure and Section V describe the approach of our system and describes the process of making machine understandable data. Section VI elaborates the process to publish geographic data of Bangladesh as Linked Open Data (LOD). Section VII includes the utilization of our proposed Linked Open Data (LOD) to experiment with SPARQL. Concluded remarks along with some future directions of our work is described in Section VIII.

II. General Terminologies

In this section we introduced to readers some basic definition of terminologies in the domain of semantic web used throughout this research paper. In this stage we familiarized Ontology, Linked Data, Resource Description Framework (RDF)[4], Uniform Resource Identifiers (URIs) and RDF Triple to comprehend the essence of our paper.

A. Ontology

In the field of semantic web, an ontology is an explicit, formal specification of a shared conceptualization of a domain of interest [17], [18]. It acts as the backbone of the semantic web vision [19], [20] which is consider as the next generation web. An ontology contains core ontology, logical mapping, knowledge base, and lexicon. Furthermore, a core ontology, S, is defined as a five tuples

$$S = (C, \leq_C, R, \sigma, \leq_R),$$

consisting of two disjoint sets C and R whose elements are called concepts and relations respectively. The notation $\leq_C$ represents a partial order on C, called concept hierarchy or "taxonomy". A partial order $\leq_R$ represents on R, called relation hierarchy,where $r_1 \leq_R r_2$ iff $dom(r_1) \leq_C dom(r_2)$ and $ran(r_1) \leq_C ran(r_2)$ and a function represented by $\sigma : R \longrightarrow C \times C$ called signature of binary relation where $\sigma(r) = \langle dom(r), ran(r) \rangle$ where $r \epsilon R$, domain $dom(r)$ and range $ran(c)$ [21].

[4]www.w3.org/RDF/

B. Linked Data

The ontology knowledge base is a structure

$$KB = (C, R, I, \imath_C, \imath_R)$$

consisting of two disjoint sets C and R as defined before, a set I whose elements are called instances, two functions $\imath_C$ and $\imath_R$ called concept instantiation and relation instantiation respectively.

With the structure of ontology and ontology knowledge base, we develop a method of publishing structured data so that it can be interlinked with diverse data sources. As a result of the interlinked data sources, both a machine or a human can explore specific data on the web is named Linked Data.

C. RDF, RDF Triple and URI

RDF is a standard model for data interchange as a triple format to interlink data on the web. In this model, resources are described using HTTP-URIs that create globally unique name without centralized management and can be easily accessing information of a resource avoiding information redundancy on the web. Moreover, an RDF triple is a piece of knowledge having subject, predicate and object. The Fig. 1 demonstrates the basic structure of triple based linked data to comprehend semantic knowledge representation.

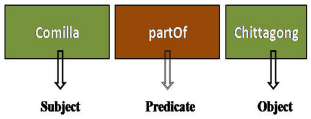


Fig. 1. RDF Triple for "Comilla $partOf$ Chittagong"

The subject of a triple is a URI, the address of the resource that we want to describe in a domain while the predicate of a triple is the description of a fact about a certain domain that represents the relationship between subject and object. We depict the RDF graph representation for $Comilla\ partOf\ Chittagong$ in Fig. 2.

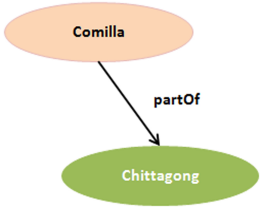


Fig. 2. RDF Graph for the Triple "Comilla $partOf$ Chittagong"

III. General Administrative Structure of Bangladesh

Bangladesh is a country of Southern Asia, officially $The\ People's\ Republic\ of\ Bangladesh$ lies on the geographical coordinates from $20^0 34N$ to $26^0 38N$ latitude and from $88^0 01E$ to $92^0 41E$ longitude. The logical relationship among different levels of the administrative structure of Bangladesh is depicted in the Fig. 3. Each level has its corresponding

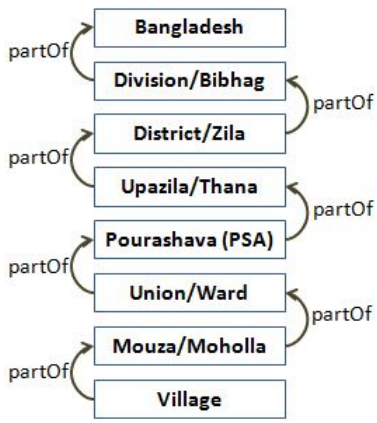


Fig. 3. Logical Relationship between Different Levels of Bangladesh

geographic information consisting their name, latitude, longitude and so on. Each level in this structure has a logical relationship with its upper and lower levels. To identify the specific geographic location of the administrative structure using *partOf* relation from the Fig. 3 in the following manner:

The country level: This entity is the root of the hierarchy represents *Bangladesh* and it is not explicitly defined in our database. This entity is assigned to the country class.

The division level: We create an entity for each division, *first administrative level* and a *partOf* relation between each division and the country. They are described in the division class.

The district level: We define each district as an entity in our domain and a *partOf* relation between each district and the division. We describe each district in the district class.

The upazila level: There is a *partOf* relation between each upazila and the district. In the metropolitan area the third level administrative structure is thana (*police station*) instead of upazila. This entity is assigned to upazila class.

The pourashava level: The pourashava level has *partOf* relation with upazila and assigned to pourashava class. The code 99 used to link with union otherwise related to wards.

The ward level: Each ward is an entity and we create a *partOf* relation between each ward and the pourashava. Each ward is assigned to a ward class.

The moholla level: We create an entity for each moholla and a *partOf* relation between each moholla and the ward. They are assigned to the moholla class. This is one of the populated place level.

The union level: Each union is a part of pourashava defining pourashava code 99. We created an entity for each union and there exist a *partOf* relation between each union and the pourashava. All unions are assigned to union class.

The Mouza level: We define each mouza as an entity and create a *partOf* relation between each mouza and the union. They are assigned to mouza class.

The village level: Villages are populated place of Bangladesh and the lower level of the administrative hierarchy. We create an entity for each village and a *partOf* relation between each village and the mouza. They are assigned to the village class.

In spite of these different levels and their relationships, data of each level is connected with another level to create a chain of data in our semantic database.

IV. Geographic Data Related to the Administrative Structure of Bangladesh

The geographic data of Bangladesh is not convenient from a single data provider's organization. It is a challenging task to collect data and validated these data using both manual and automatic process from different sources. The geographic designations to name and address of each location originated by Bangladesh Bureau of Statistics [5] is available in Microsoft Excel file, *GeocodeBD.xls*. This file contains 7 divisions, 64 districts, 500 upazilas and 509 administrative thanas, 265 pourashava or municipalities, 2407 wards, 4451 union council, 67100 mouza or moholla and 87968 villages. Among these data 96,600 populated places are obtainable. The fragment of data in *GeocodeBD.xls* is shown in the Fig. 4. Therefore

Division: CHITTAGONG District/Zila: COMILLA
Upazila/Thana: BARURA Ward/Union: DAKSHIN KHOSBAS
Mouza/Moholla: BARA HORIPU

Division Code	Zila Code	Upazila/Thana Code	PSA	Union Code	Mouza/Moholla Code	Village Code	Area Name
20	19	09	99	69	081		BARA HORIPUR
20	19	09	99	69	081	01	BARA HORIPUR
20	19	09	99	69	081	02	SRIRAMPUR

Fig. 4. Fragment of Data in the *GeocodeBD.xls* file

another file published as the features file, *feature.xls* contains information about geographical classes such as Division, District, Upazila and so on. The data snippet of *feature.xls* is given in the Fig. 5. Moreover, it is a formidable task to

Feature Class	Feature Codes	Feature	Feature Description	Feature Description
BA	BA00	Division	a primary administrative division of a country, such as a division of Bangladesh	একটি দেশের প্রথম প্রশাসনিক বিভাগ সমূহ, যেমন বাংলাদেশের ক্ষেত্রে বিভাগসমূহ
BA	BA01	District / Zila	a subdivision of a first-order administrative division, such as district or zila of Bangladesh	প্রথম প্রশাসনিক বিভাগ সমূহের উপবিভাগ সমূহ, যেমন বাংলাদেশের ক্ষেত্রে জেলা সমূহ
BA	BA02	Upazila / Thana	a subdivision of a second-order administrative division, such as upazila or thana of Bangladesh	দ্বিতীয় প্রশাসনিক বিভাগ সমূহের উপবিভাগ সমূহ, যেমন বাংলাদেশের ক্ষেত্রে উপজেলা বা থানা সমূহ

Fig. 5. Data snippet of *feature.xls* file

aggregate latitude and longitude values for each location. To

[5] http://bbs.gov.bd/RptGeoCode.aspx

Geonames_id	Name	Alternate_Name	Latitude	Longitude	Feature_Class	Feature_Code	Admin1_Code	Population	Elevation
6416752	Aaartali	Aaartali	21.38761	92.22421	P	PPL	84	40519	2
6414576	Adar Char	Adar Char	21.99498	92.15439	P	PPL	84	0	3
1213248	Adhunagar	Adhunagar	21.98239	92.07578	P	PPL	84	0	5

Fig. 6. Data segment of *BD.xls* file

do this task we conscientiously congregate this latitude and longitude coordinate values from the geo-names geographical database *BD.xls* [6]. The data segment of this file is given in the Fig. 6. In this data source, there are 53719 records are serviceable for geographic locations. We extract latitude and longitude values only for administrative structure related entities of Bangladesh. In spite of these data source, there are a huge number of challenges to incorporate appropriate values with *Geocode-BD* for a particular location. In this database the feature code ADM1 is equivalent to our division in *Geocode-BD*. Therefore feature code ADM2 is also similar to district or zila. There are 68 districts are available in this dataset instead of 64 districts from the dataset provided by the government of Bangladesh. In order to identify duplicate records for Jhalokati, Khagrachhari and Pirojpur districts are removed to aggregate latitude and longitude values. Moreover, another record for parbattya chattagram district is not a valid second level administrative entity of Bangladesh. Furthermore, in the ADM3 level has 308 data records in the geo-names geographic server. However, there are only four classified records for the fourth administrative level and 248 records missing in *BD.xls* file for ADM3 level. In order to identify these missing data and challenges of automatic extract coordinate values of an entity we successfully complete this task by google geo-coding api. Moreover, each geographic entity is represented with latitude and longitude coordinates in Cartesian WGS84 (World Geodetic System 1984) format, a standard coordinate reference system mainly used in cartography, geodesy and navigation to represent geographical coordinates on the Earth [7].

V. Our Approach

The approach of our application is to classify resources and instances of the system including their respective features. The subsequent phase of our research is to recognize concepts and relationship types of resources for making machine understandable data repository coined a name $Geo-Bangladesh$, *a semantic database related to the geographic information of Bangladesh* from available data those are related to our domain of interest. In the third stage, we published this knowledge repository as Linked Open Data (LOD) to create link with related web of contents to address the interoperability issues on the web. In the consecutive step, a lot of experiment is performed to retrieve and inference specific information using SPARQL [8] from our machine understandable data repository. The overview of our approach is depicted in Fig. 7 to comprehend the essence of the system.

[6]http://download.geonames.org/export/dump/
[7]https://www1.nga.mil/ProductsServices/
[8]http://www.w3.org/TR/rdf-sparql-query

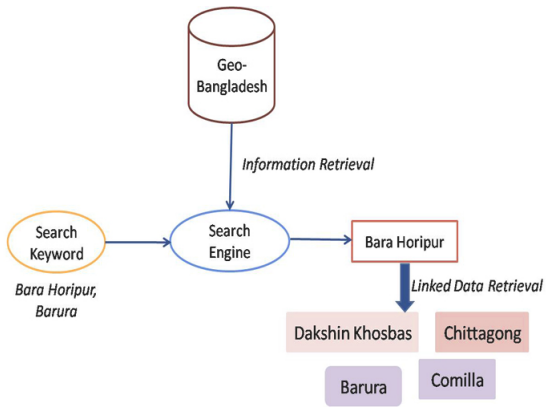


Fig. 7. Overview of Our Information Retrieval System

A. Making Data Machine Understandable

An ontology defines common vocabularies for researchers [22] who need to share information of a domain on the web of data. It is more easier to publish data in RDF using available vocabularies from the web. In the case of unavailability of vocabularies, users can propose for a suitable vocabulary to describe domain data on the web. In order to identify these vocabularies we develop $Geo-inspired$ ontology to describe our domain knowledge explicitly. The basic steps to develop an ontology is [22]: defining classes and arranging these classes in a taxonomic (subclass-superclass) hierarchy, defining relationships with other classes called slots and allowed values for these slots. However, our knowledge data repository $Geo-Bangladesh$ contains fundamental geographic data related to the administrative structure taking into account classes, sub-classes, instances, relations, object property, data property and general axioms etc. It is important to select a vocabulary to maximize interoperability with wider consensus on the web. We use Dublin Core (DC) ($dublincore.org/documents/dces/$) and DCMI-BOX ($http://dublincore.org/documents/dcmi-box/$) standard vocabularies to encode geographic meta-data of our domain. For example, a resource type is defined in a class using $http://www.w3.org/1999/02/22-rdf-syntax-ns\#type$ vocabulary. Each instance of a class consist their related features such as name and comment is published using $http://www.w3.org/2000/01/rdf-schema\#label$ and $http://www.w3.org/2000/01/rdf-schema\#comment$ vocabularies. Moreover, latitude and longitude of each location are defined using common vocabularies of $http://www.w3.org/2003/01/geo/wgs84_pos\#lat$ and $http://www.w3.org/2003/01/geo/wgs84_pos\#long$. Therefore, the geographic bounding box attributes for north, south, east and west are mapped to $dcmibox:northlimit$, $dcmibox:southlimit$, $dcmibox:eastlimit$ and $dcmibox:westlimit$ respectively. Therefore, each resource of our domain is linked

Subject	Predicate	Object
<http://www.skeim.org#20>	<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>	<http://www.skeim.org#division>.
<http://www.skeim.org#20>	<http://www.w3.org/2000/01/rdf-schema#label>	Chittagong@en.
<http://www.skeim.org#20>	<http://www.w3.org/2000/01/rdf-schema#comment>	"Information about Chittagong division@en".
<http://www.skeim.org#20>	<http://www.w3.org/2003/01/geo/wgs84_pos#lat>	"22.330391@en".
<http://www.skeim.org#20>	<http://www.w3.org/2003/01/geo/wgs84_pos#long>	"91.82518000000004@en".
<http://www.skeim.org#20>	<http://www.w3.org/2000/01/rdf-schema#partOf>	<http://www.skeim.org#Bangladesh>.
<http://www.skeim.org#2019>	<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>	<http://www.skeim.org#district>.
<http://www.skeim.org#2019>	<http://www.w3.org/2000/01/rdf-schema#label>	"Comilla@en".
<http://www.skeim.org#2019>	<http://www.w3.org/2000/01/rdf-schema#comment>	"Information about Comilla district @en".
<http://www.skeim.org#2019>	<http://www.w3.org/2003/01/geo/wgs84_pos#lat>	"23.455959@en".
<http://www.skeim.org#2019>	<http://www.w3.org/2003/01/geo/wgs84_pos#long>	"91.18203689999996@en".
<http://www.skeim.org#2019>	<http://www.w3.org/2000/01/rdf-schema#partOf>	<http://www.skeim.org#20>.

Fig. 8. Machine Understandable Geographic Data of Bangladesh in N-TRIPLE format

with other resources using RDF schema definition $http://www.w3.org/2000/01/rdf-schema\#partOf$. The Fig. 8 gives us a close look of the machine understandable representation of geographic data in RDF format. The representation of these data create link of a resource with other resources to form a directed, labeled graph which is portrayed in the Fig. 9.

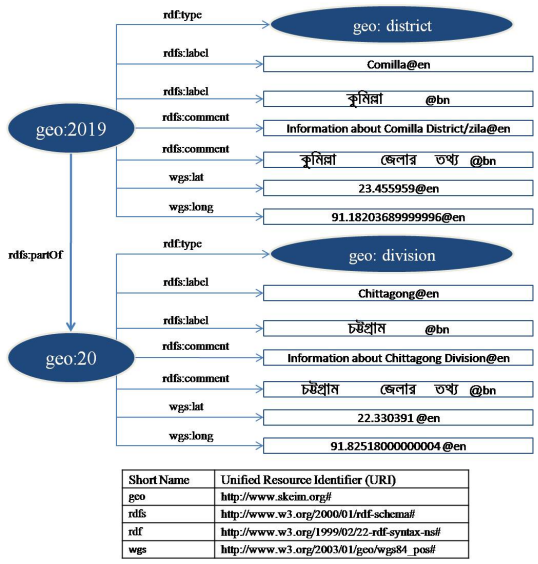


Fig. 9. Machine Understandable Geographic Data of Bangladesh in a Graph

VI. Published Geographic Data on Bangladesh as Linked Open Data

Linked data focuses on the ontological level and inferences to create knowledge base interoperable global data sets. In this technology we published structured data in RDF using HTTP-URIs for public access and hence widely called as Linked Open Data (LOD). The LOD of our domain provides reusable data for other applications on the web and retrieve specific geographic data related to the administrative structure of Bangladesh using semantic search engine. The inventor of the World Wide Web Tim Berners-Lee outlines four principles of Linked Data [23] as:

1) Use URIs as names for things
2) Use dereference-able URIs (HTTP-URIs) so that people can look up those names
3) When someone looks up a URI, provide useful information, using the standards (RDF, SPARQL)
4) Links to resources in other datasets should be included in order to enable the discovery of more data.

Adhering to these four linked data principles we published $Geo - Bangladesh$ to make data interconnected with one more data repositories, $Bangladeshi - Citizen$ to achieve the interoperability issue. In this case, resource define villages in the $Geo - Bangladesh$ data source is shared with $Bangladeshi - Citizen$ to retrieve the complete address of a citizen.

VII. Knowledge Representations and Reasoning

Our proposed application share resources or instances from $Geo - Bangladesh$ without defining correlative data in other applications. Chaining different web content increases the number of knowledge-intensive task that can be carried out automatically [24]. The semantic web supports technology for describing, discovering and accessing web applications that can be formed a chain of interoperable web applications dynamically [25]. Moreover, we use java based semantic web application development framework, *Jena* for rule based inference and SPARQL, a query language for semantic data. Furthermore, we consider the following knowledge pieces, *S-1, S-2, ...,S-7* with their associations to identify each entity of the administrative structure of Bangladesh and derive important information based on isa and $partOf$ relation.

S-1: Bangladesh ***isa*** country
S-2: Bara Horipur ***isa*** village and ***partOf*** Bara Horipur Mouza
S-3: Bara Horipur ***isa*** Mouza and ***partOf*** Dakshin Khosbas union
S-4: Dakshin Khosbas ***isa*** union and ***partOf*** Barura Upazila
S-5: Barura ***isa*** Upazila and ***partOf*** Comilla zila/district
S-6: Comilla ***isa*** district and ***partOf*** Chittagong division
S-7: Chittagong ***isa*** division and ***partOf*** Bangladesh

The inference system can derive both direct and indirect relation from *Geo-Bangladesh*, using their relationship between the subject and the object of knowledge pieces. This inference engine also generate many more necessary statements by applying various rules on the available existing

triples in our semantic data source. Therefore the experiment for the information of all district names those are more closest in terms of their distance from the district where Bara Horipur village exist and similar experiments also performed for other entities of Bangladesh i.e, division, upazila, union, mouza and moholla and vice-versa. The specific hierarchical data retrieval procedure is as follows:

Step-1: Based on *part-of* relation retrieve subject or object for an entity.
Step-2: Apply subject or object from step-1 to retrieve other data.
Step-3: Repeat step-1 and step-2 until the specific goal achieved.

Moreover, the semantic query to retrieve URIs of all division from our knowledge-base data source $Geo-Bangladesh$ using SPARQL semantic search engine is portrayed in the Fig. 10 and their respective result in the Fig. 11.

```
PREFIX rdf: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX rdfs: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX Geo-Bangladesh: < http://www.skeim.org #>
SELECT ?URI
 WHERE
   {
       ?URI rdf:type Geo-Bangladesh:division
       };
```

Fig. 10. Semantic query using SPARQL search engine

URI
http://www.skeim.org/10
http://www.skeim.org/20
http://www.skeim.org/30
http://www.skeim.org/40
http://www.skeim.org/50
http://www.skeim.org/55
http://www.skeim.org/60

Fig. 11. Semantic query Result for URIs of all division

VIII. Conclusion

Now-a-days machine understandable geographic data of a country is essential to decrease data profusion for other applications those use these compatible data on the web. $Geo - Bangladesh$, a geographic semantic data repository of Bangladesh reduces the formation of similar data for other applications. This machine understandable web content is interrelated with available semantic data sources to address data interoperability issues in a faster and more efficient way. RDF query is performed to retrieve and inference more specific information that reflects the effectiveness of our application. Using $Geo-Bangladesh$, we extend this research to map our locality to ease search ability by mass people who have access to mobile technology.

References

[1] J. Chomicki and P. Revesz, "Constraint-based interoperability of spatiotemporal databases," *Geoinformatica*, vol. 3, no. 3, pp. 211–243, 1999.

[2] Z. Peng, "A proposed framework for feature-level geospatial data sharing: a case study for transportation network data," *International Journal of Geographical Information Science*, vol. 19, no. 4, pp. 459–481, 2005.

[3] C. Zhang and W. Li, "The roles of web feature and web map services in real-time geospatial data sharing for time-critical applications," *Cartography and Geographic Information Science*, vol. 32, no. 4, pp. 269–283, 2005.

[4] Y. Bishr, "Overcoming the semantic and other barriers to gis interoperability," *International Journal of Geographical Information Science*, vol. 12, no. 4, pp. 299–314, 1998.

[5] F. Harvey, W. Kuhn, H. Pundt, Y. Bishr, and C. Riedemann, "Semantic interoperability: A central issue for sharing geographic information," *The Annals of Regional Science*, vol. 33, no. 2, pp. 213–232, 1999.

[6] C. Jones, H. Alani, and D. Tudhope, "Geographical information retrieval with ontologies of place," *Spatial Information Theory*, pp. 322–335, 2001.

[7] A. Cohn, "The challenge of qualitative spatial reasoning," *ACM Computing Surveys*, vol. 27, no. 3, pp. 323–325, 1995.

[8] B. Smith and D. Mark, "Geographical categories: an ontological investigation," *International Journal of Geographical Information Science*, vol. 15, no. 7, pp. 591–612, 2001.

[9] A. Frank and D. Mark, "Language issues for gis," *Geographical information systems: Principles and applications*, vol. 1, pp. 147–163, 1991.

[10] O. Consortium *et al.*, "Geography markup language (gml) 3.0," *Open GIS Implementation Specification,[Online]. Disponível: http://www. opengis. org/docs/02-023r4. pdf*, 2001.

[11] P. Vretanos, "Web feature service implementation specification," *Open Geospatial Consortium Specification*, pp. 04–094, 2005.

[12] J. de La Beaujardière, "Ogc web map service interface, version 1.3. 0," *Open Geospatial Consortium*, 2004.

[13] J. Brodeur, Y. Bedard, G. Edwards, and B. Moulin, "Revisiting the concept of geospatial data interoperability within the scope of human communication processes," *Transactions in GIS*, vol. 7, no. 2, pp. 243–265, 2003.

[14] F. Fonseca, M. Egenhofer, P. Agouris, and G. Câmara, "Using ontologies for integrated geographic information systems," *Transactions in GIS*, vol. 6, no. 3, pp. 231–257, 2002.

[15] D. Mark, B. Smith, and B. Tversky, "Ontology and geographic objects: An empirical study of cognitive categorization," *Spatial Information Theory. Cognitive and Computational Foundations of Geographic Information Science*, pp. 747–747, 1999.

[16] T. Berners-Lee, J. Hendler, O. Lassila *et al.*, "The semantic web," *Scientific american*, vol. 284, no. 5, pp. 28–37, 2001.

[17] T. Gruber *et al.*, "Toward principles for the design of ontologies used for knowledge sharing," *International journal of human computer studies*, vol. 43, no. 5, pp. 907–928, 1995.

[18] R. Studer, V. Benjamins, and D. Fensel, "Knowledge engineering: principles and methods," *Data & knowledge engineering*, vol. 25, no. 1, pp. 161–197, 1998.

[19] T. Berners-Lee and M. Fischetti, *Weaving the Web: The original design and ultimate destiny of the World Wide Web by its inventor*. DIANE Publishing Company, 2001.

[20] A. Maedche and S. Staab, "Ontology learning for the semantic web," *Intelligent Systems, IEEE*, vol. 16, no. 2, pp. 72–79, 2001.

[21] M. Ehrig, *Ontology alignment: bridging the semantic gap*. Springer, 2006, vol. 4.

[22] N. Noy, D. McGuinness *et al.*, "Ontology development 101: A guide to creating your first ontology," 2001.

[23] C. Bizer, T. Heath, and T. Berners-Lee, "Linked data-the story so far," *International Journal on Semantic Web and Information Systems (IJSWIS)*, vol. 5, no. 3, pp. 1–22, 2009.

[24] Y. Ding, D. Fensel, M. Klein, and B. Omelayenko, "The semantic web: yet another hip?" *Data & Knowledge Engineering*, vol. 41, no. 2, pp. 205–227, 2002.

[25] M. Daconta, L. Obrst, and K. Smith, "The semantic web: a guide to the future of xml, web services, and knowledge management," 2003.

Breast Cancer Classification from Ultrasonic Images Based on Sparse Representation by Exploiting Redundancy

Abdullah Al Helal*, Khawza I Ahmed*, Md. Saifur Rahman† and S. Kaisar Alam‡
*Department of Electrical and Electronic Engineering
United International University (UIU), Dhaka 1209, Bangladesh
Email: {helal, khawza}@eee.uiu.ac.bd
†Department of Electrical and Electronic Engineering
Bangladesh University of Engineering and Technology (BUET), Dhaka 1000, Bangladesh
Email: saifur@eee.buet.ac.bd
‡Rutgers University, Piscataway, NJ, USA
Email: kaisar.alam@ieee.org

***Abstract*—We present a Sparse Representation-based Classifier (SRC) that provides superior performance in terms of high Area Under the Receiver Operating Characteristic (ROC) Curve (AUC) in classifying benign and malignant breast lesions captured in ultrasound images. Although such a classifier was proposed for face recognition, it has been proposed in medical diagnosis from ultrasonic images in this work for the first time. The classifier is based on ℓ_1-norm based sparse representation of a patient's test data in terms of linear combination of the features of the benign and malignant test lesions available in the training set. The proposed classifier introduces an index called Sparsity Rank (SR) for the classification obtained from the normalized energy of the weights as a linear combination of the global sparse representation of the ultrasound images of the training set. The performance of the classifier is further enhanced to a great extent by two ways: first, by intelligently combining the features extracted from the multiple ultrasound scan of the same mass, and the second, by using the optimal feature set obtained by a suboptimal strategy that avoids the time exhaustive brute force approach that has a combinatorial search space. With all the enhancements an AUC of 0.9802 has been achieved, when training and testing sets are chosen by leave-one-out approach from the data set.**

I. Introduction

The average life expectancy has been extended to 80 years in developed countries and around 60 years in developing countries [1]. Due to the advancement in medical science and engineering, the primary cause of death have been shifted toward the cardiovascular diseases and cancers from the earlier reasons of airborne and direct-contact infectious diseases [1], [2]. Among various cancers, breast cancer is the second most fatal one among women after lung cancer [2]. However, early detection of it reduces the risk of death to a great extent.

Ultrasound image or mammogram of the breast can effectively be used for early detection of cancer by Computer Aided Diagnosis (CAD). The motivation behind the use of CAD is to reduce or eliminate the dependency on the subjective judgement of the radiologist on the detection and the diagnosis of the breast lesion (benign or malignant). This can only happen when CAD becomes consistent in always giving the correct diagnosis decision.

For an ideal classifier the ROC has 100% Area Under the ROC Curve (AUC). However, a real classifier's AUC falls below that. One logical question that arises is where the operating point on ROC should be chosen. This depends on the application, e.g., medical diagnosis or radar detection, related with the ROC. In cancer diagnosis in the breast, the operating point needs to be chosen so that true positive rate (TPR) or *sensitivity* is nearly 100% even at the cost of high false positive rate (FPR) or $1-$*specificity* and thus the overall accuracy. In this paper, we present an enhanced classifier based on Sparse Representation-based Classifier (SRC) that provides us near-to-ideal operating points in the ROC curve with an AUC of 0.9802.

A basic difference between the problem of classifying a biometric class, e.g., face recognition, and that of classifying a disease group, e.g., tumor classification, is the nature of the classes themselves. For example, each class in face recognition refers to a particular person, whose data samples, i.e., images inherently contain a good deal of overlapping information resulting in redundancy. On the contrary, each class in tumor classification refers to a particular tumor type, but not a particular patient or even a particular tumor, thus de-emphasizing the effect of inclusion of redundancy. Specifically, a test sample in face recognition is ideally to be matched against the training samples of the same person, whilst a test sample in tumor classification is ideally to be matched against the training samples of the same type rather than the same tumor. Repeated scanning of the same lesion may enable us to exploit the inherent redundancy of the data by providing a complementary view.

Since CAD tools allow for investigating abnormal features that may otherwise be invisible on the screening image, several groups have worked on developing CAD tools to aid in

978-1-4799-3497-3/13/$31.00 ©2013 IEEE

automating the diagnosis of breast masses. Shankar et. al. utilized non-Rayleigh statistics of the Radio Frequency (RF) envelope to achieve an AUC of 0.94 [3]. Joo et. al. implemented an artificial neural network technique using shape features to get an AUC of 0.95 [4]. Chang et. al. used texture features in a Support Vector Machine (SVM) approach, reporting a Positive Predictive Value (PPV) of 89.74 and a Negative Predictive Value (NPV) of 96.24 [5]. Huang and Chen implemented SVM to classify breast tumors with an AUC of 0.97 [6]. Bocchi et. al. used a video clip of ultrasound to come up with an accuracy up to 0.97 [7]. Yanjiao et. al. used SVM on boundary-based features obtaining an accuracy of 0.95 [8]. Tan et. al. used coronal spiculation from automated 3-D breast ultrasound images to obtain an AUC of 0.93 [9]. This list is not exhaustive. A good number of work did not consider the effect of the ultrasound scanning system, hence did not compensate for the contribution of the system.

Sparse representation is a recently developed classification technique based on the non-conventional compressive sensing principle. This representation falls into the category of reconstructive methods. Different reconstructive and discriminative methods have been cited in the literature for breast cancer classification. While discriminative methods aim at maximizing the separation of signal classes, and thus are, in general, capable of classifying ideal signals more efficiently than reconstructive methods [10], [11], their performance degrades in presence of noise and outliers due to their nonrobust nature [12]. Sparse representation is shown to achieve state-of-the-art performance in image denoising and face recognition [13]. The fact that ultrasound is inherently a noisy process motivates us to investigate sparse representation in classifying cancer images.

The prime contributions of this paper are as follows. SRC as a classifier is introduced for the first time in the problem of medical diagnosis. The performance of SRC in face recognition has proven to yield high accuracy [14], where the test image can usually be expressed as the linear combination one of the classes, i.e., persons, in the database and thus sparsity lies in all the other classes with respect to the matched class. However, in classification of tumor, as explained earlier, it may not always be possible to express the test tumor of unknown class in terms of the linear combination of the training tumors of one of the two classes. This may result in non-sparsity and eventual failure of the correct detection. Also, the degree of sparsity may become poor as only two classes, i.e., benign and malignant are considered. This limitation of SRC is overcome when the information of multiple scans of the same tumors is combined by the averaging of the SR during testing and afterwards the decision is made comparing the average SR with the set threshold in the classifier.

This paper has been inspired from and is the result of [15]. To the best of our knowledge, none has so far reported a combining technique of multiple scans that results in superior performance, i.e., AUC as high as 0.98, although works are reported [16] where the features available from multiple scans of a same tumor are averaged before used during training and testing. To improve the performance of SRC further, an optimization strategy for the feature set to be used in the classifier is proposed that searches in a reduced space compared to the computationally prohibitive combinatorial search space.

The rest of the paper is organized as follows. Section II formulates the problem statement in a formal manner. Section III presents the proposed SRC based method adopted in classifying cancerous from the benign tumors. The exploitation of redundancy provided by the multiple scans that provides superior performance is elucidated in Section IV. Section V presents the optimization of the feature set. The experiments and results are explained and discussed in Section VI followed by our conclusion in Section VII.

II. Problem Statement

The task of tumor classification translates to the problem of correctly assigning every test tumor sample to its benign or malignant status using labeled training tumor samples. The working database includes RF ultrasonogram scans of patients. Each image in the database is denoted by $\mathscr{I} = (\mathbf{P}, R)$ where $\mathbf{P}$ is a two-dimensional (2D) grid of pixels (x, y) and $R(x, y)\colon (x, y) \in \mathbf{P}$ is the image intensity function that corresponds to the RF signal received through ultrasonogram scans. The lesion $\mathbf{L} \subset \mathbf{P}$, visible in the RF image, is traced and separated using a trained human guided boundary tracing program, so that $\mathbf{L}$ defines the set of all pixels within the boundary of the lesion. Thus, we will identify an ultrasonogram tumor scan $\mathscr{L} = (\mathbf{L}, R)$ into one of the c classes based on the available training samples. In our work, we have identified the tumor to have either benign or malignant masses, thus implying $c = 2$. We could say, for example, benign tumors form class 1 and malignant tumors form class 2.

III. Methodology

The salient steps of the proposed method are outlined in Figure 1. The steps are explained in the following.

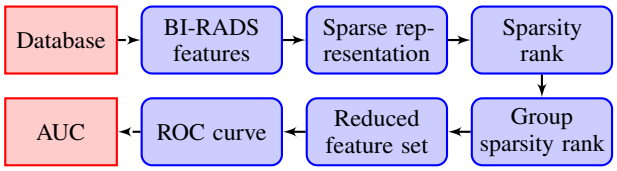


Fig. 1. Flowchart depicting the classification procedure.

A. Feature Extraction

The American College of Radiology (ACR) has developed Breast Imaging Reporting and Data System (BI-RADS) to classify breast masses [17], [18]. In this work, we use quantitative descriptors to simulate the BI-RADS features so as to make our evaluation reproducible and to reduce inter-observer variability.

B. Test Samples as Sparse Linear Combinations of the Training Samples

Having generated the feature set, we can now interpret any test tumor sample $(\mathbf{L}, R)$ as a projection on a *feature space* represented by a vector $\mathbf{b} \in \mathbb{R}^m$. We incorporate a simple but efficient model which assumes that the samples from a single class lie on a linear subspace.

Thus given sufficient number of training samples n_i, arranged in a matrix form,

$$\mathbf{A}_i = [\mathbf{b}_{i,1}, \mathbf{b}_{i,2}, \ldots, \mathbf{b}_{i,n_i}] \tag{1}$$

from class i, any test sample $\mathbf{b}$ from this class can be written as

$$\mathbf{b} = a_{i,1}\mathbf{b}_{i,1} + a_{i,2}\mathbf{b}_{i,2} + \cdots + a_{i,n_i}\mathbf{b}_{i,n_i} \tag{2}$$

for $a_{i,j} \in \mathbb{R},\ j = 1, 2, \ldots, n_i$.

As the membership i of the test sample is not known during training, we could use a large matrix consisting of all n training samples

$$\mathbf{A} = [\mathbf{A}_1 | \mathbf{A}_2 | \ldots | \mathbf{A}_c] \tag{3}$$

with columns $\mathbf{b}_\lambda$, where $\{\lambda \in \Lambda_i\}$ is a set of indices that corresponds to training samples $\mathbf{A}_i$ from class i, enabling us to write the equation set

$$\mathbf{b} = a_1\mathbf{b}_1 + a_2\mathbf{b}_2 + \cdots + a_n\mathbf{b}_n \tag{4}$$

or,

$$\mathbf{b} = \mathbf{A}\mathbf{x}_r \tag{5}$$

where

$$\mathbf{x}_r = [a_1, a_2, \ldots, a_n] \tag{6}$$

with only non-zero entries a_λ, $\lambda \in \Lambda_i$ associated with class i.

What makes the apparently obvious and standard approach of solving the linear system in (5) difficult is that the solution is not unique dictated by the fact that m and n need not be equal due to the independence between the training sample size and the number of features extracted from the samples.

If $m > n$, i.e., the number of features surpasses the number of training samples, the system is overdetermined, there may or may not be a unique solution. However, considering the usual simplified modeling and measurement error, methods exist to find an 'approximate' solution to the system, and these methods usually find the correct solution $\mathbf{x}_r$.

On the other hand, if $m < n$, the system is underdetermined and there is no unique solution. To make things worse, test samples, when corrupted with noise, may not satisfy (5) even in case of an overdetermined system. The problem of non-uniqueness of the solution can be dealt with looking for an Mean Square Error (MSE), or equivalently, ℓ_2-norm solution:

$$\mathbf{x}_2 = \arg\min_{\mathbf{x}} \|\mathbf{x}\|_2, \quad \text{subject to} \quad \mathbf{A}\mathbf{x} = \mathbf{b}. \tag{7}$$

Not only does this solution not equal the correct solution $\mathbf{x}_r$, it generally is dense in that it contains non-zero entries in almost coordinates, making it ineffective classifying the sample.

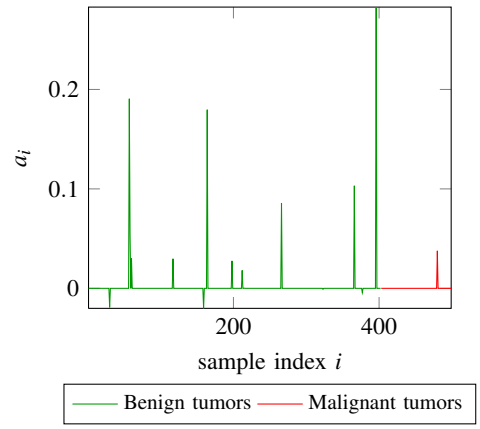


Fig. 2. A typical simulation result showing the contribution of the training samples to build up the test sample. Samples from index 1 to index 402 correspond to the actual class. Some of the coefficients of the other samples are small but non-zero.

The model, as represented in (6) of $\mathbf{x}_r$, dictates that any test sample $\mathbf{b}$ from class i can be expressed in terms of the training samples from that class. This representation is *sparse* in the sense that at most n_i entries are non-zero for an ideal case. Intuitively, the more terms that are zero, the more accurately a test sample can be associated with its native class; thus we seek the sparsest solution:

$$\mathbf{x}_0 = \arg\min_{\mathbf{x}} \|\mathbf{x}\|_0, \quad \text{subject to} \quad \mathbf{A}\mathbf{x} = \mathbf{b}, \tag{8}$$

where $\|\cdot\|_0$ is the ℓ_0-norm that counts the non-zero elements of a vector. Although theoretically this sounds simple, solving (8) is NP-hard [19], which requires a practical algorithm. It has been shown that if $\mathbf{x}_r$ is sparse enough, it can solved for by solving a totally different program

$$\mathbf{x}_1 = \arg\min_{\mathbf{x}} \|\mathbf{x}\|_1, \quad \text{subject to} \quad \mathbf{A}\mathbf{x} = \mathbf{b}, \tag{9}$$

making the ℓ_1-norm solution equal to the ℓ_0-norm solution [20].

C. Ranking and Classification of the Test Samples

Ideally, training samples from only the true class of a test sample will contribute to the sparse representation of the test sample. Thus, in the sparse representation of a test sample, coefficients from a single class will show up as non-zero. We can then detect this class as membership of the test sample under consideration:

$$\text{detection}(\mathbf{b}) = \arg\text{any}\{\lambda \in \Lambda_i \colon a_\lambda \neq 0\}. \tag{10}$$

In practical cases, however, noise and measurement errors may cause small, but non-zero coefficients associated with the wrong class. Figure 2 displays a typical sparse representation of a benign test sample where samples from position 1 to position 402 correspond to the actual class and other positions correspond to the wrong class. We could design a classification scheme engaging the global sparse representation that can take advantage of the subspace structure of each class. For example, ℓ_1-norm of the values of a single class can provide information

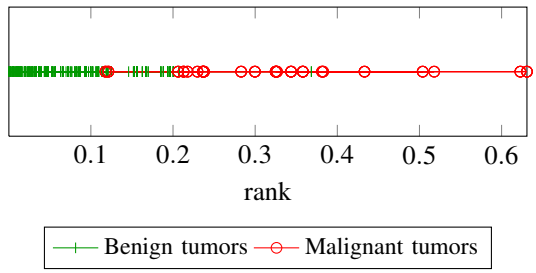


Fig. 3. A typical simulation result showing the distribution of the rank of each class.

about how the whole class contributes to the solution. The problem with this choice is that ℓ_1-norm does not discriminate between small (hence not significant) values over large (hence significant) values much as norms of higher order (or any of their strictly monotonic functions), e.g., ℓ_2-norm, do.

Thus, we introduce a ranking function Sparsity Rank (SR) (or simply, rank), which assigns a score or rank to each solution measuring how close it is to be a 'perfect' member of the class 2:

$$\text{rank}(\mathbf{b}) = \frac{\|\{a_\lambda : \lambda \in \Lambda_2\}\|_2}{\|\mathbf{x}\|_2}. \tag{11}$$

The normalization introduced in (11) is found to cancel out the effect of non-uniform class sizes. Ideally all class-1 samples will have a rank of 0 and all class-2 samples a rank of 1. In practical cases, recovery error $\mathbf{x}_r - \mathbf{x}_1$ equals $\{a_\lambda : \lambda \in \Lambda_2\}$ for class 1, and $\{a_\lambda : \lambda \in \Lambda_1\}$ for class 2, so that a rank will identify a sample through its proximity to one of the two extremities, and it would be enough to compare the rank of a sample with a threshold rank_{th} to determine its class:

$$\begin{aligned} \text{if } \ \text{rank}(\mathbf{b}) &> \text{rank}_{\text{th}} \\ \text{detection}(\mathbf{b}) &= 2, \\ \text{else } \ \text{detection}(\mathbf{b}) &= 1. \end{aligned} \tag{12}$$

The performance of a classifier for a particular threshold can be measured using *accuracy*, and a paired value of TPR and FPR, among other measures such as precision and F-measure.

Figure 3 shows the spread of the rank of each class for a typical run of the classification method. We note one interesting thing about ranking the samples (as opposed to classifying outright) as follows. Classification at this stage requires that all class-2 samples precede a predetermined threshold value for a perfect classification. On the contrary, ranking the samples requires that all class-2 samples only precede all class-1 samples. In this setup, the user can later choose a suitable threshold that separates the two classes.

A straightforward implication of using ranking is that we can vary the threshold which essentially selects the operating point of classification and thus gives different classifiers. Flexibility in choosing the operating point, e.g., for a particular TPR is desired in some biomedical applications. For instance, legal issues about diagnosing a patient dictates that a radiologist/oncologist/physician would prefer a classification scheme with a high TPR over the one with a low TPR, even at the the price of a high FPR.

Varying the threshold over all possible values of rank gives us $(FPR_\lambda, TPR_\lambda)$ for each b_λ. The set $\{(FPR_\lambda, TPR_\lambda)\}$, when plotted on the xy plane, is called the ROC curve. While each point on the curve corresponds to an operating point accuracy (or other measures), the overall performance of the classifier can thus be measured by the AUC.

IV. Exploitation of Redundancy

From the point of view of information theory, redundancy in a piece of information can be used to recover it more efficiently or check for any corruption, possibly by noise. This motivates us to build a database by including redundant labeled data, which can be exploited for better classification. In sonography, we can scan a tumor several times, e.g., from different angles. In face recognition, this can mean taking photos of the same person with different facial expressions and lighting conditions.

The key point of including redundancy is that the internal structure of data be the same. Of course, it would carry no further information if the identical data is obtained several times. If we consider a measurement error, the same data will be read differently due to the measurement error:

$$\mathbf{b} = \mathbf{A}\mathbf{x}_r + \mathbf{e}, \tag{13}$$

$\mathbf{e}$ being the noise contributing to the measurement error. Furthermore, sonography itself is a noisy process because it not only measures tissue variations but also includes speckles [21]. For example, for fully developed speckle, the Signal to Noise Ratio (SNR) is only 1.91.

The tendency of the prevalent noise patterns such as Gaussian noise to have a zero mean motivates us to device a mechanism to act on the measurement error which will be propagated to the source of the noise to reduce it as well.

Under the assumption that the noise involved in (13) is zero-mean, averaging of the rank enables us to head for a zero-mean noise through *Optimal Sparsity Rank*, or simply optimal rank

$$\text{rank}_{\text{opt}}(\mathbf{b}^\theta) = \mathscr{A}[\text{rank}(\mathbf{b}^{\theta_j})] \tag{14}$$

where $\{\mathbf{b}^{\theta_j}\}$, $j = 1, 2, \ldots, N^\theta$ belong to the samples or 'realizations' of θ-th 'unique' one out of n^θ test samples and $\mathscr{A}[\cdot]$ indicates averaging over j.

V. Optimization of Feature Set

It should be noted here that the projection of data in lower subspaces can lead to improved classification. For example, we can seek for reducing an m-dimensional feature vector into a lower-dimensional vector such that this new vector represents the most effective feature set for this classification problem. This fact warrants the search for appropriate feature reduction. This can be related to the principle of shortest data description in model selection which specifies the most compact representation of a model to be preferred for classification [22]. One additional benefit of reducing the number of features is the reduced computational cost associated with low-dimension data.

The BI-RADS dictionary suggested by ACR describes a set of features of ultrasound images of breast tumors which can potentially improve the accuracy of classification, hence diagnosis. This work implements quantitative descriptors of the mentioned features, which is traditionally assessed subjectively by physicians, to remove dependence on the operator and the expertise of the physicians.

To assess the relative significance of the BI-RADS features against one another, an exhaustive search can be employed that attempts to find the most significant features contributing to the correct separation of classes. In particular, for $\mathbf{b} = \{ f_k \}$, $k = 1$, 2, …, m, every possible subspaces of $\mathbf{b}$, i.e., every possible subsets of $\{ f_k \}$ can be substituted for $\mathbf{b}$ to search for the subspace that is responsible for the maximum AUC. This would be the optimum result considering the inclusion or exclusion of each feature in the feature space. If the computational cost associated with the exhaustive search is impractical, as found to be the case with the current work, a suboptimal search can be implemented which seeks recursively for a feature set by initializing from $\{ f_k \}$ and deciding upon exclusion of each exclusion of each feature if the exclusion improves the resulting AUC and then going for next feature until the result reaches a saturation. The outcome will be suboptimal since choice of the sequence of features will affect the result. Consequently, the optimum solution may not be reached.

VI. Experiment and Results

A. ATL Database

We tested our classification algorithm on a data set containing RF data and biopsy diagnosis for $n = 502$ tumor scans out of which $n_1 = 405$ were benign and $n_2 = 95$ malignant. Many of the tumor lesions were scanned at multiple scan planes, N^θ ranging from 0 to 25, providing complementary data about the tumors. A total of $n^\theta = 130$ unique tumors, 104 benign and 26 malignant, were scanned. An expert sonographer or radiologist examined the mammographically detected masses using an Advanced Technology Laboratories (ATL) (now Philips) Ultrasound UM-9 HDI scanner with a 7.5 MHz L10-5 linear-array transducer. Standard breast examination procedures were followed. The operator set the transducer at a constant power level and a constant transmit focal length. A Spectrasonics Inc. data acquisition module allowed digital acquisition of RF echo-signal data, at a sampling rate of 20 MHz with a resolution of 14 bits. A set of analysis regions corresponding to the tumors was traced and set as Region of Interest (ROI).

Calibration was done to account for the three contributions other than tissue information: the combined two-way transfer function of the transducer and the data acquisition module, the two-way range-dependent diffraction function raised from beam properties, and the acoustic attenuation.

Spectral analysis was performed, and the signal over a bandwidth of 6 dB was considered for the linear regression during the analysis of the spectrum: the parameters of interest were the midpoint of the regression line (M), its intercept at zero frequency (I), and its slope (s). Estimation of envelope of the RF data for each position of the transducer generates the envelope data.

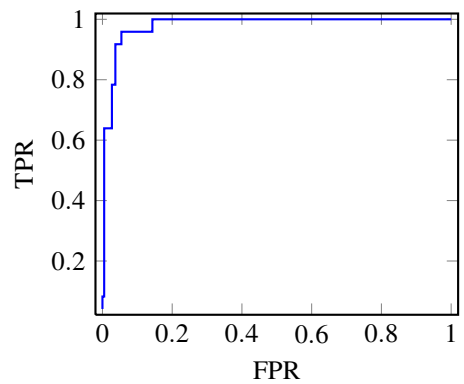


Fig. 4. Classification with suboptimal BI-RADS feature set using the optimal rank.

The traced tumors vary from pixel size 111×192 to 265×192, which have been downsampled to a general length of 25 for each of the image types. The length has been chosen as to allow performing the simulation in a reasonable time.

Experiments have been conducted to justify how the proposed classifier performs and how much each step signifies. The experiments start with a bare implementation of the SRC, and then include the subsequent steps shown in Figure 1. The following subsections describe the implementations of the steps, though the details of the outcomes and their interpretations have been omitted, and will be reported in a journal article.

B. The Role of Feature Extraction

BI-RADS features are reported to allow for better classification results. We have implemented 25 different, but not necessarily independent, features from the BI-RADS criteria and their variants. These are: echogenicity, heterogeneity, co-occurrence contrast, co-occurrence contrast with image in dB, 4-neighborhood-pixels algorithm to measure texture, 4-neighborhood-pixels algorithm with image in dB, Hurst coefficient to measure heterogeneity, Hurst coefficient with image in dB, two different implementations of shadowing, two different implementations of relative absorption, margin definition (gradient-based), margin definition (edge-based), lesion aspect ratio, lesion area, lesion aspect ratio divided by area, lesion compactness, lesion roundness, lesion convexity, lesion solidity, lesion form factor, lesion Kolmogorov dimension to measure border irregularity, lesion Minkowski dimension to measure border irregularity, and lesion Hausdorf dimension to measure border irregularity. Description of the features is outside the scope of this paper. However, these are described in details in [16].

Implementation of BI-RADS features and using the rank to classify can be shown to result in a vast improvement over the performance on the classification using images alone without using BI-RADS features. In particular, the features give highly discriminatory information necessary to separate the two classes.

C. Exploiting Redundancy

Most tumors of the ATL database have several scans at different scanning planes. We employ rank_{opt} instead of rank on the database to exploit the duplication of data.

D. Optimizing Feature Set

We employ the proposed recursive search to come up with a suboptimal set of features. Using this suboptimal feature set instead of the original feature set of 25 features results in the ROC curve, as shown in Figure 4, with an AUC of 0.9802.

The suboptimal set includes 15 features, namely, echogenicity, heterogeneity, co-occurrence contrast, 4-neighborhood-pixels algorithm to measure texture, 4-neighborhood-pixels algorithm with image in dB, Hurst coefficient to measure heterogeneity, Hurst coefficient with image in dB, shadowing, relative absorption, lesion aspect ratio, lesion area, lesion compactness, lesion roundness, lesion form factor, and lesion Hausdorf dimension to measure border irregularity.

VII. Conclusion

We have introduced SRC for medical diagnosis from ultrasonic images. Our work demonstrates that SRC is effective in discriminating between malignant and benign lesions when the SRs of multiple scans are averaged and feature sets are optimized. Therefore, a judicial use of multiple scans indeed makes a difference and it is always a better option compared to the complicated biopsy when the latter one is less convenient and less comfortable for a patient and also costlier in several order of magnitudes. In the future, we will investigate how the multiple scans can be combined intelligently for better performance.

References

[1] W. M. Saltzman, *Biomedical Engineering: Bridging Medicine and Technology (Cambridge Texts in Biomedical Engineering)*. Cambridge University Press, Jun. 2012.

[2] A. C. Society, "Cancer facts & figures 2013," Atlanta: American Cancer Society, Annual Report, 2013.

[3] P. M. Shankar, V. A. Dumane, C. W. Piccoli, J. M. Reid, F. Forsberg, and B. B. Goldberg, "Computer-aided classification of breast masses in ultrasonic B-scans using a multiparameter approach," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 50, no. 8, pp. 1002–1009, 2003.

[4] S. Joo, Y. S. Yang, W. K. Moon, and H. C. Kim, "Computer-aided diagnosis of solid breast nodules: Use of an artificial neural network based on multiple sonographic features," *IEEE Transactions on Medical Imaging*, vol. 23, no. 10, pp. 1292–1300, 2004.

[5] R.-F. Chang, C.-J. Chen, M.-F. Ho, D.-R. Chen, and W. K. Moon, "Breast ultrasound image classification using fractal analysis." Taichung: IEEE, 2004, pp. 100–107.

[6] Y.-L. Huang and D.-R. Chen, "Support vector machines in sonography: Application to decision making in the diagnosis of breast cancer," *Clinical Imaging*, vol. 29, no. 3, pp. 179–184, 2005.

[7] L. Bocchi, F. Gritti, C. Manfredi, E. Giannotti, and J. Nori, "Semiautomated breast cancer classification from ultrasound video." Barcelona: IEEE, 2012, pp. 1112–1115.

[8] Z. Yanjiao, L. Jiangli, C. Ke, and P. Yulan, "Boundary-based feature extraction and recognition of breast tumors using support vector machine," vol. 3, Chengdu, 2009, pp. 89–92.

[9] T. Tan, B. Platel, H. Huisman, C. I. Sánchez, R. Mus, and N. Karssemeijer, "Computer-aided lesion diagnosis in automated 3-D breast ultrasound using coronal spiculation," *IEEE Transactions on Medical Imaging*, vol. 31, no. 5, pp. 1034–1042, 2012.

[10] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711–720, 1997.

[11] A. M. Martinez and A. C. Kak, "PCA versus LDA," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 2, pp. 228–233, 2001.

[12] S. Fidler, D. Skočaj, and A. Leonardis, "Combining reconstructive and discriminative subspace methods for robust classification and regression by subsampling," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 3, pp. 337–350, 2006.

[13] M. Elad and M. Aharon, "Image denoising via learned dictionaries and sparse representation," in *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, vol. 1. IEEE, 2006, pp. 895–900.

[14] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 210–227, 2009.

[15] A. A. Helal, "Breast cancer classification from ultrasonic images based on sparse representation," Master's thesis, Bangladesh University of Engineering and Technology, 2013.

[16] S. K. Alam, E. J. Feleppa, M. Rondeau, A. Kalisz, and B. S. Garra, "Ultrasonic multi-feature analysis procedure for computer-aided diagnosis of solid breast lesions," *Ultrasonic Imaging*, vol. 33, no. 1, pp. 17–38, 2011.

[17] A. C. of Radiology, Ed., *Breast Imaging Reporting and Data System (BI-RADS) Atlas*. Reston, VA: American College of Radiology, 2003.

[18] E. B. Mendelson, W. A. Berg, and C. R. Merritt, "Toward a standardized breast ultrasound lexicon, BI-RADS: Ultrasound," *Seminars in Roentgenology*, vol. 36, no. 3, pp. 217–225, 2001.

[19] E. Amaldi and V. Kann, "On the approximability of minimizing nonzero variables or unsatisfied relations in linear systems," *Theoretical Computer Science*, vol. 209, no. 1–2, pp. 237–260, 1998.

[20] D. L. Donoho, "For most large underdetermined systems of linear equations the minimal l1-norm solution is also the sparsest solution," *Communications on Pure and Applied Mathematics*, vol. 59, no. 6, pp. 797–829, 2006.

[21] M. Halliwell, "A tutorial on ultrasonic physics and imaging techniques," *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine*, vol. 224, no. 2, pp. 127–142, 2010.

[22] A. d'Aspremont, L. El Ghaoui, M. I. Jordan, and G. R. Lanckriet, "A direct formulation for sparse PCA using semidefinite programming," *SIAM review*, vol. 49, no. 3, pp. 434–448, 2007.

Automatic Slice Growing Method Based 3D Reconstruction of Liver with its Vessels

Md. Zahangir Alom, Moin Mostakim
Dept. of Computer Science and Engineering
BRAC University, Dhaka, Bangladesh
{zahangir, mostakim}@bracu.ac.bd

Rubel Biswas, Amitabha Chakrabarty
Dept. of Computer Science and Engineering
BRAC University, Dhaka, Bangladesh
{rubelbiswas, amitabha}@bracu.ac.bd

Abstract—**In the recent years, reconstructing 3D liver and its vessels from abdominal CT volume images becomes an inevitable and necessary research field. In this paper, a method of 3D reconstruction of liver with its vessels has been implemented, which involves volume preprocessing, de-noising, segmentation, contouring, and combination of different modalities. An advanced liver segmentation algorithms have been proposed: the first one is a 2.5D method that utilizes automatic Slice Growing Method (SGM) to segment liver part of each slice of a data set. It takes advantage of curvature control of level set segmentation method to distinguish liver and adjacent organs. It is proved that the result of this proposed method is much better than simple 3D level set method in liver segmentation. In the case of liver vessel segmentation, we have proposed an improved smoothing method dedicate to 3D vascular volume which results from region growing segmentation method. The cooperation of region growing method and proposed smoothing method has been demonstrated the possibility of efficient vessel segmentation with very accurate results. And the results indicate that our method is suitable for anatomical studying and surgical planning.**

***Index Terms*— 3D level set, SGM, Vessel Segmentation, Liver Segmentation, MRI, CT**

I. Introduction

The influence and impact of digital images on modern society is tremendous, and digital image processing is now a critical component in science and technology. With the fast computers and signal processors available in the 2000s, digital image processing has become the most common form of image processing, and is generally used because it is not only the most versatile method, but also the most economic. Digital image processing allows us to analyze, enhance, classify, project, and extract features of interest in images to gain more useful information by various techniques and algorithms, such as linear filtering [1], Principal Components Analysis (PCA) [2], Hidden Markov Models (HMM) [3], neural networks [4], partial differential equations [5] and wavelets [6]. It is widely used in the field of computer vision, face detection, feature detection, medical image processing, microscope image processing, morphological image processing, remote sensing and more [7][8][9]. Medical imaging is a technique which is used to create and process medical digital images of the human body (or parts and function thereof) for clinical purposes or medical science. The techniques such as computerized tomography (CT), magnetic resonance imaging (MRI), or positron emission tomography (PET) are nowadays the standard instruments for computer-aided diagnosis. The rapid progress in computerized medical image reconstruction, and the associated developments in analysis methods, has propelled medical imaging into one of the most important sub-fields in scientific imaging [10] [11] [12].In the medical community, DICOM (Digital Imaging and Communications in Medicine) files are widely used for interchange of images and other data, and it is also a standard for handling, storing, printing, and transmitting information in medical imaging [13]. A DICOM file usually contains scalar data elements to represent 2D grey level image. In this paper, 3D level set and ASG algorithm have been applied for segmenting liver with its vessel. Moreover, the performance analysis of this ASG against 3D level set algorithm has been represented in the result and discussion section

The rest of this paper is organized as follows. Section II presents on related works of medical imaging, liver and vessel segmentation, proposed system implementation is discussed in Section III. Experimental results and performance evaluation are given in Section IV. Conclusion and future direction of this work is discussed in Section V.

II. Related works

More recently, several approaches have been tested in the MICCAI workshop on segmentation challenge. Some automatic liver segmentation approaches have been proposed by few researchers. For example, Chi et al. [18] have presented an automatic strategic active contour method integrated a rotational template matching, and k-means clustering followed by rib cage area local edge enhancement, with a gradient vector flow geometric snake. Susomboon et al. [19] proposed a hybrid approach, which consists of four stages, for automatic liver segmentation in 2D data. In [20], the authors developed automatic 3D liver segmentation method based on a combination of a constrained free-form and statistical deformable model. However, all of these automatic methods cannot handle our organ-confused CT images, and the accuracy of them is not good for every data set. To improve the accuracy of liver segmentation, few interactive methods have been proposed by researchers. For example, Reinhard et al. [17] proposed an interactive refinement approach to 3D liver segmentation based on a 3D refinement

concept utilizing a hybrid desktop/virtual reality user interface. Slagmolen et al. [15] used an atlas in combination with non-rigid registration to do 2D liver segmentation from CT images. In [21], the authors evaluated a semi-automatic liver segmentation method based on a level-set approach and a dynamically adapted speed function. Some of these methods using costly interactive device that cannot easily generalize, and some methods just focused on 2D Liver segmentation. Freiman et al. [22] have classified vessel segmentation approaches into six categories: 1) histogram based methods, 2) semi-automatic region growing based methods; 3) topology based methods; 4) level-set based methods; and 5) geometrical model based methods; 6) hybrid approaches. The latest category of liver segmentation was presented by Esneault et al. [23]. They provided an automatic method for liver vessel segmentation, however, their final result as shown in Figure 1 was very inaccurate and most of veins were lost. In [24], Erdt et al. have proposed GPU-based hepatic vessel segmentation from CT data. It includes vessel enhancement and vessel segmentation, which were processed on graphical processing units (GPU). However, review of 3D vessel segmentation techniques has been well summarized by D. Lesage et al.[25] And a little bit older review is published by C. Kirbas and F. Quek [26] [27].

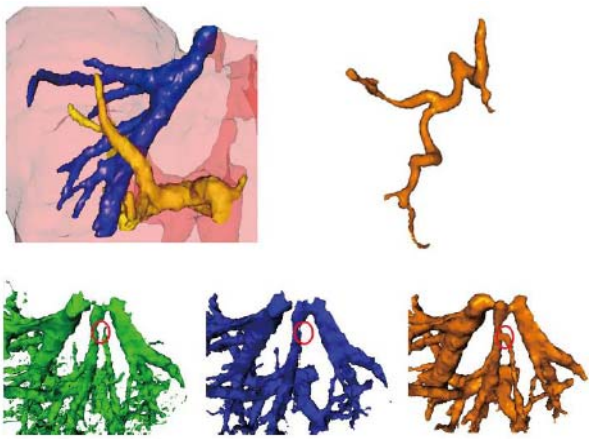


Fig.1. Results of liver segmentation were presented in 2008

There are many applications dedicated to medical image processing and visualization such as ImageJ [28], 3D Slicer [29], ParaView [30], ITK-Snap [31] etc. Especially, popular applications that are able to perform 3D segmentation of liver and vessel are 3D Slicer and ITK-Snap. ITK-Snap is also a free software application. It is used to segment structures in 3D medical images with semi-automatic segmentation methods. However, confused parameters in some algorithm such as level-set methods need professional knowledge. According to free software survey, there is no public software dedicated to the segmentation and visualization of liver with its vessels. Even if some software's can meet a few of our ideas, they do not provide more supports for our further requirements, such as liver volume calculation and vessel length calculation. Thus these factors lead us to develop a software application for our research.

III. Proposed System Implementation

Primary characteristics of medical images are spatial resolution, contrast resolution, linearity, noise and artifacts which affect image quality in CT. There have been several algorithms reported for image denoising, and each approach has its assumptions, advantages, and limitations [33]. In this paper, a few of image denoising algorithms mentioned Curvature Flow Denoising algorithm (CFD has been used. For liver segmentation, a more complicated and time-consuming denoising algorithm, Curvature Anisotropic Diffusion Denoising (CADD), has been implemented relies on the work of Whitaker et al. CAAD introduced a partial differential equation for denoising images, and is demonstrably superior to anisotropic diffusion for denoising images that are approximately piecewise constant.

A. Segmentation Methods

Image segmentation is the task of splitting a digital image into one or more regions of interest. In this paper, image segmentation methods are mainly used and discussed in medical imaging context. By investigating segmentation techniques in the state of the art, popular medical image segmentation approaches can be classified into seven categories: (1) thresholding methods, (2) region growing methods, (3) deformable models, (4) level-set approaches, (5) Fuzzy connectedness, (6) watershed method, and (7) atlas-guided approaches[14][19].In this study, region growing method is used for vessel segmentation since the intensity values of vessels from CT image are basically able to be distinguished. While threshold guided level set algorithm is used in liver involved segmentation include SGM liver segmentation algorithm. This algorithm is 2.5D segmentation algorithm with different purpose.

B. Region Growing

Region growing segmentation method can be classified as one of the pixel based image segmentation algorithms because it involves the selection of an initial seed points. Region growing methods can quickly separate the regions that have the same properties, such as a range of intensity value in grey level images. Starting with a single pixel/voxel in a 2D/3D image, region growing method will check the neighboring pixel/voxel and add them to the region if they are satisfied some conditions such as average intensity, size, shape, and motion. And all pixels/voxels that are satisfied the condition will be repeat previous steps to explorer new neighbors. However, for liver vessel segmentation, threshold based region growing method is a sophisticated method that ensures both efficiency and accuracy.

C. Level Set

The level set method, since introduced by Osher and Sethian has seen widespread application in image processing (segmentation), computer graphics (surface reconstructions) and physical simulation (particularly fluid simulation). It is numerical techniques for tracking the evolution of contours and surfaces. Instead of manipulating the contour directly, the

contour is embedded as the zero level set of a higher dimensional function called the level-set function, $\Phi(X,t)$. . The level-set function is then evolved under the control of a differential equation. At any time, the evolving contour can be obtained by extracting the zero level-set, shown in equation 1, from the output.

$$\Gamma\big((X),t\big) = \{\Phi(X,t) = 0\}$$

The main advantages of using level sets is that arbitrarily complex shapes can be modeled and topological changes such as merging and splitting are handled implicitly, as shown in Fig. 2.

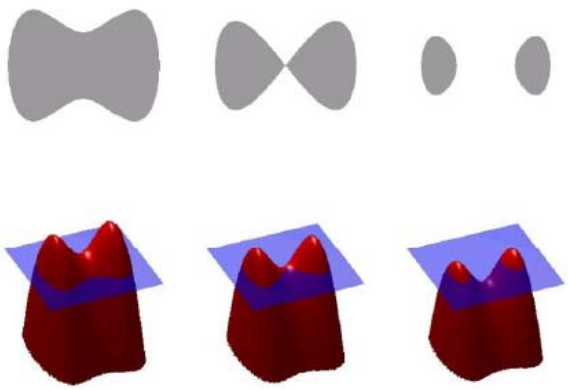


Fig.2. The relationship between level-set function (top) and contour (bottom).

The evolution of the contour or surface is governed by a level set equation. The solution tended to by this partial differential equation is computed iteratively by updating Φ at each time interval. The general from of the level set equation is

$$\frac{\partial \Phi}{\partial t} = -|\nabla \Phi| \cdot F,$$

where is the velocity term that describes the level set evolution. By manipulating, level set will be guided to different areas or shapes. Level set algorithm will constrain curvature of an organ and stay clear from confusing organs that have similar intensity values. Taking advantage of shape detection feature of level-set algorithm, the roughly circular orbit like liver shape can be extracted precisely. However, these kinds of approaches are time-consuming especially in 3D volume segmentation. In this experiment, a threshold guided 2.5D level-set algorithm has been developed in the consideration of speed and other necessary advantages. It works by constructing a speed function which is close to zero at the upper and lower bounds of an intensity window, effectively locking the propagating front onto those edges.

D. Liver Reconstruction

Two advanced 2.5D liver segmentation methods have been proposed in this thesis. The first one is the SGM that takes advantage of level-set algorithm to handle 2D image segmentation, and according to previous segmented slices to estimate segmentation conditions for adjacent slices. The other method uses the shape similarity between adjacent slices and mapping segmented result in previous slice to current slice and start segmentation from the mapped result. Experiments prove the latter method is largely improved speed and accuracy of liver segmentation.

E. Slice Growing Method (SGM)

To get accurate 3D liver model, a five-stage approach has been proposed in this section. Firstly, CT images should be preprocessed using edge-preserved smoothing algorithm to reduce image noise without losing topological information. Secondly, a selected seed slice in the middle part of liver volume is applied to a segmentation method to accurately get liver part in it. The next step is to get the contour of the segmented liver in the seed slice and estimate an appropriate seed point position for adjacent slices to perform afterward segmentation procedure. And then each slice will be processed by analogy with previous segmented slices and conditions. Finally, 3D liver surface can be reconstructed by all of these segmented slices.

F. Seed Slice Selection and Point generation algorithm

For liver related research, DICOM images scanned from CT devices are specially focused on liver part. In general, liver part will be definitely appeared in the middle slices of whole volume The seed slice is one of the images contains liver part of a data set, and it is manually selected from the middle slices. Start from this seed slice, each slice will be processed sequentially by two working threads in opposite directions as shown in Fig.3.

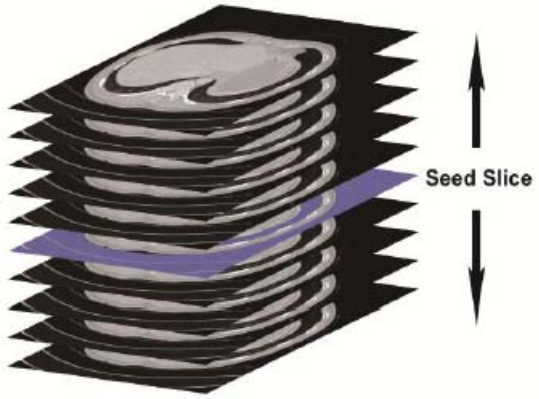


Fig.3. Two processing directions from the seed slice

The whole volume now is divided into two parts, and each part will be sent to a working thread to perform slice growing segmentation. There are two basic inputs, which are intensity thresholds and seed point, in adopted 2D level-set algorithm. Intensity thresholds consist of lower and upper values can be defined according to mean intensity value with its standard deviation. Seed points for each slice can be obtained by proposed automatic seed generation algorithm described in the next section. The complete seed selection procedure is described as below and Fig.4.

Input: CT Volume and segmented seed slice
Output: Segmented binary volume
1. Get the contour of segmented liver shape
2. Calculate distance between the seed point and each contour point
3. Get the minimum distance and the corresponding point (say M')
4. Using the seed point as center point, get the symmetric point of M', say S'
5. Using seed point and S' as diameter, make the first circle
6. Make several rings by gradually halving the diameter
7. Build seed lists according to ring number
8. Jump to next adjacent slice
9. For each seed position list (start from the list of the innermost ring)

i. Choose random seed point
ii. Perform segmentation to working image with it
iii. Calculate shape mutation
iv. If shape mutation is reasonable
Break;
Else
Continue;
10. Goto 1

According to Step 3, 4 and 5, all of these seed rings will be definitely located inside the liver, as shown in Figure 5. One of the advantages of using multiply seed rings is to maintain seed positions not shifted much between adjacent slices. At the same time, a program has more choice to get suitable seeds if previous segmentations are failed. It means the quality of seeds in inner rings is superior to the outer rings.

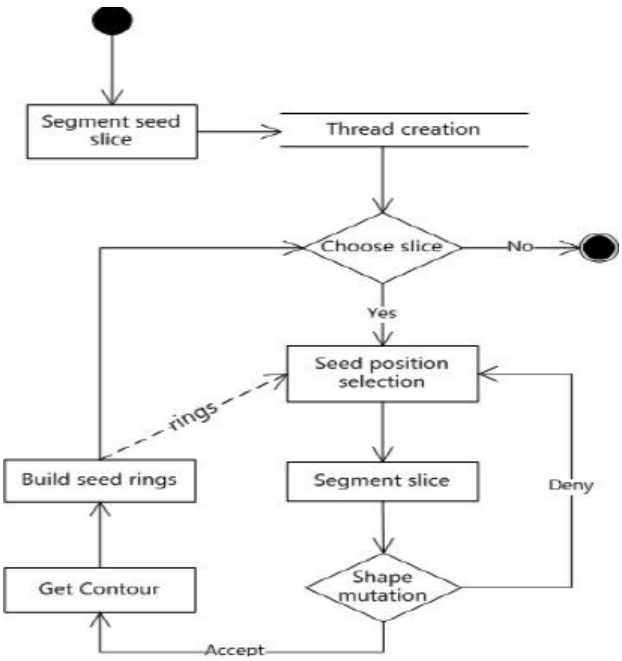


Fig. 4. Main segmentation procedure of working threads

In the Step 9, the algorithm will calculate the shape variation between current processing slice and its reference slices to reduce the risk of deflection. The reference slices are defined as adjacent segmented slices as shown in Fig.5. If the segmented liver shape changed too drastically referring to reference slices, it has to choose another seed point for it. The detection of shape mutation is to compare the shape of current segmented slice with reference slices. If the contour voxel number of liver shape of current segmented slice isn't changed according to the variation tendency of reference slices, the segmented result would not be accepted. And then the program will try segmenting again until the segmentation result is acceptable.

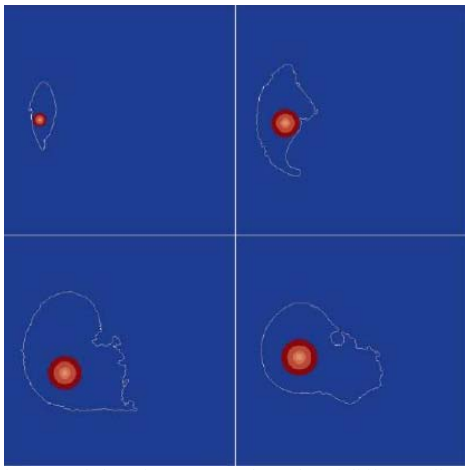


Fig.5. Seed rings in segmented results of different slices.

G. Vessel Reconstruction

In this research, vessels including hepatic vein, port vein and artery are extracted by region growing method discussed in section X, and the output volume is binary volume, as we can see in Fig. 6, which is not smooth enough as the output of using level set algorithm. An Improved Aliasing Aritfacts Removing (IAAR) algorithm is proposed to make vessel surface smoother without losing vessel information

Fig. 6. The result of vessel segmentation by modified region growing method.

Whitaker (Whitaker, 2000) has implemented a solution of reducing aliasing artifacts in binary data based on a narrow band level set method. It recovers grey-scale embeddings from binary data by introducing a regularization that picks the

feasible solution with the least surface area. However, vessels in the segmented 3D volume are fine tubular which only occupy few voxels, some vessels thus turned disappeared and disconnected after applying the Aliasing Aritfacts Removing (AAR) algorithm as shown in the left of Fig.7.

Fig. 7. The surface of applying IAAR algorithm

To solve this problem, we expand edge-voxels according to its vector field of normal to make the model one voxel larger than original model before applying AAR, as shown in Fig. 8.

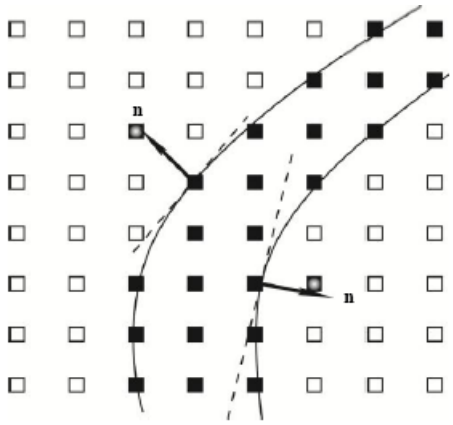


Fig. 8. Voxel expanding according to normal direction.

In the edge of the volume, each voxel will be expanded along with its normal n. The intensity value of nearest voxel (the gradient fills in Fig. 8) in the direction will be marked into the segmented volume. Finally the vessel will be enlarged with a reasonable range.

IV. Experiments and Performance Analysis

The experimental results and performance of this proposed system has been discussed.

A. Environmental setup

CT images to be used in this research were provided by Medical School of Chonbuk National University, and the software is developing on 2.53 GHz Dual Core E7200@2.53GHz with 4GB RAM with Windows 7 x64 installed. Experimental CT images are high resolution images, which are 512x512 with 1mm interval as shown in Fig. 9. Intensity values inside a liver are very similar to the main artery and heart.

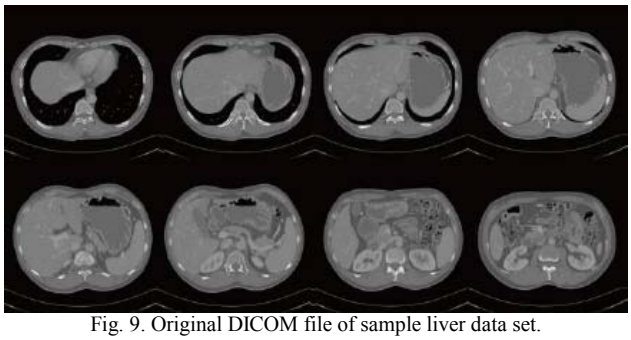


Fig. 9. Original DICOM file of sample liver data set.

B. Liver Segmentation

The combination of segmented liver and its vessels are shown in Fig. 10.

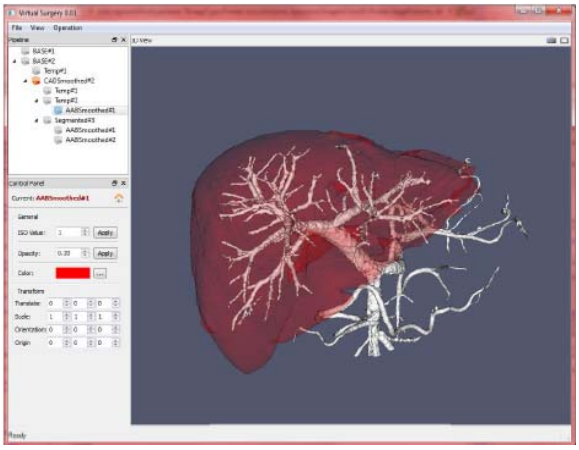


Fig. 10. Final combination of liver and vessels in Virtual Surgery application.

C. Slice Growing Method (SGM)

Segmented results of each slice were then transferred to display thread in order to observe real time display during the processing of both 3D level set and SGM, and it can be also manually interrupted and resume at any time. After reconstructed all of these segmented slices, 3D liver volume and its surface generated from Slice Growing Liver Segmentation method are shown in Fig. 11.

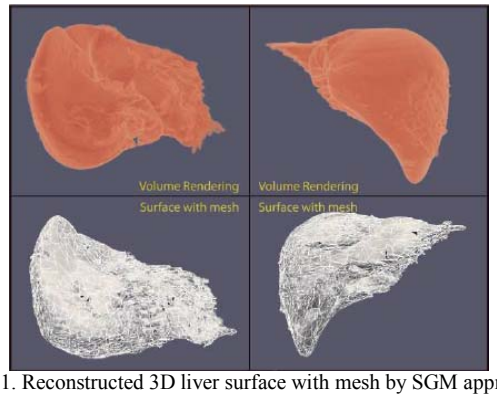


Fig. 11. Reconstructed 3D liver surface with mesh by SGM approach.

The top two pictures in Figure 11 are using direct volume rendering technique, and the other two pictures are 3D surface with meshes. As we can see from Figure 11, 2D segmented

results are seamlessly reconstructed to 3D volume. After checking segmented 2D results with original images, no slice has been lost and each segmented result is quite precise. Some of the slices have been processed more than once for different reasons. For example, seed point positions obtained from reference slices may not be suitable for current slice because there were many small vessels inside the liver, and these positions sometimes were taken by these vessels. Another important factor result in multiple-processing, which means a slice has been processed more than once, is the shape mutation between two adjacent slices caused by the complex curvature variation of liver edge, in this case, proposed algorithm will choose another seed point position for better segmentation result. The multiple-processing of the same slice will spend longer time. We have tried some data sets to demonstrate the practicality of a proposed approach, and the extra processing times of experimental data-sets are summarized in Table I.

TABLE I. MULTIPLE PROCESSING OF LIVER IMAGES

Liver Volume	Volume Resolution	Slice that include liver	Processing times	Multiple-processing
1	512x512x424	144	158	9.7%
2	512x512x464	216	233	7.8%
3	512x512x245	245	261	6.5%
4	512x512x458	165	178	7.8%
5	512x512x188	188	201	6.9%
6	512x512x142	142	158	11.3%

Since an image will probably be processed more than once by the SGM algorithm, the number of processed time of a volume is sensitive. However, even if SGM will bring multiple-processing for some slices, the processing time is still much fewer than 3D level set segmentation. Moreover, 3D level set takes larger memory usage, and while SGM just requires very low memory usage since the slice by slice processing approach will efficiently free unnecessary memory as required.

D. Vessel Segmentation

For a 512x512x245 volume with 'short' type, proposed vessel segmentation plus IAAR smoothing method were taken around 10 seconds running on experimental platform. And the segmentation results of portal vein are shown in Figure 12.

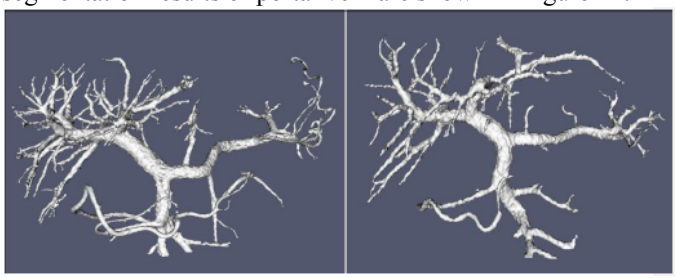


Figure 12: Two results of Vessel Segmentation with IAAR smoothed.

Compared with current published results, our segmentation result should be better on both accuracy and vision.

E. Performance Analysis

The comparison based on computational time of SGM and 3D level set shown in Table 02 and Figure 13,

TABLE II. PERFORMANCE EVALUATION TABLE

Liver Volume	**Volume Resolution**	**Processing time of 3D Level set**	**Processing time of SGM**
1	512x512x424	58.2m	726.3s
2	512x512x464	64.4m	734.8s
3	512x512x245	39.5m	656.6s
4	512x512x458	65.1m	715.4s
5	512x512x188	33.9m	433.5s
6	512x512x142	28.2m	385.9s

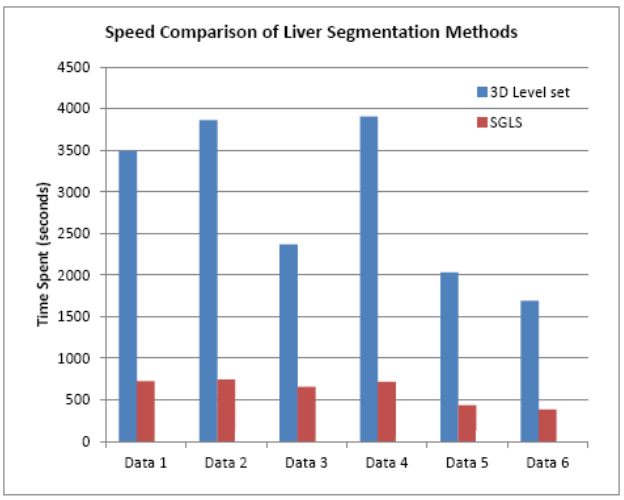


Figure 13: Efficiency comparison of proposed algorithm for liver segmentation

V. CONCLUSION AND FUTURE WORKS

A. Conclusions

3D liver segmentation is still a challenging work because of exist difficulties mentioned in Section 1. There are two big problems in 3D liver segmentation, accuracy and efficiency. Both of these two problems have been solved by proposed liver segmentation methods SGM. And the result of SGM is better than 3D Level Set. However, SGM as the prototype of our reconstruction concept and framework should be mentioned. The main idea of this proposed method is to find out accurate liver parts in each slice firstly, and then to reconstruct a whole liver surface from these segmented slices. In SGM, there are two important steps for segmenting 2D liver image. The first step is to obtain a correct seed point for each slice, and then to apply slice segmentation algorithm with the seed point. The other step is to ensure segmented liver image is accurate, that's because segmented results sometimes are not very perfect even if seed point is correct, moreover, the

segmentation algorithm will be sensitive to adjacent voxel information around a seed point. Therefore, the efficiency comparison table clearly shows that the proposed SGM algorithm worked about four times faster than 3D level set algorithm.

B. Future Works

In future, our work will be focused on improving segmentation result by applying CMM. In some data sets, the results that generated by CMM bring some unwanted noise, such as parts of heart, so a model based method that to remove other organs is a crucial point for liver segmentation. Moreover, various data sets should be tested by our software to make it stable and adoptable.

REFERENCES

[1] A. Zverev, *Handbook of filter synthesis*, Wiley New York, 1967.

[2] H. Barrett, K. Myers, and S. Dhurjaty, *Foundations of image science*, 2005.

[3] J. Li and R. Gray, *Image segmentation and compression using hidden Markov models*, Kluwer Academic Publishers, 2000.

[4] T. Lindblad and J. Kinser, *Image processing using pulse-coupled neural networks*, Springer Verlag, 2005.

[5] G. Aubert and P. Kornprobst, *Mathematical problems in image processing: partial differential equations and the calculus of variations*, Springer-Verlag New York Inc, 2006.

[6] L. Prasad and S. Iyengar, *Wavelet analysis with applications to image processing*, CRC, 1997.

[7] Rafael C. González and R.E. Woods, *Digital image processing*, Prentice Hall, 2008.

[8] J.C. Russ, *The Image Processing Handbook*, CRC Press, 2007.

[9] B. Jähne, *Digital image processing*, Springer, 2005.

[10] V. Sherrow, *Medical Imaging*, Marshall Cavendish, 2006.

[11] J. Udupa and G. Herman, *3D imaging in medicine*, CRC, 2000.

[12] L. Beolchi and M. Kuhn, *Medical imaging: analysis of multimodality 2D/3D images*, Ios Pr Inc, 1995.

[13] O. Pianykh, *Digital imaging and communications in medicine (DICOM): a practical introduction and survival guide*, Springer Verlag, 2008.

[14] P. Campadelli and E. Casiraghi, "Liver segmentation from CT scans: A survey," *LECTURE NOTES IN COMPUTER SCIENCE*, vol. 4578, 2007, p. 520.

[15] P. Slagmolen, A. Elen, D. Seghers, D. Loeckx, F. Maes, and K. Haustermans, "Atlas based liver segmentation using non-rigid registration with a B-spline transformation model," 2007, pp. 197-206.

[16] A. Wimmer, G. Soza, and J. Hornegger, "Two-stage Semi-automatic Organ Segmentation Framework using Radial Basis Functions and Level Sets," 2007, pp. 179-188.

[17] R. Beichel, C. Bauer, A. Bornik, E. Sorantin, and H. Bischof, "Liver Segmentation in CT Data: A Segmentation Re nement Approach," *Methods*, 2007, pp. 235-245.

[18] Y. Chi and F. Bello, "A Discussion on the Evaluation of A New Automatic Liver Volume Segmentation Method for Speci ed CT Image Datasets," *Heart*, 2007, pp. 167-175.

[19] R. Susomboon, D.S. Raicu, and J. Furst, "A Hybrid Approach for Liver Segmentation," *Most*, vol. i, 2007, pp. 151-160.

[20] D. Kainm, T. Lange, and H. Lamecker, "Shape Constrained Automatic Segmentation of the Liver based on a Heuristic Intensity Model," 2007, pp. 109-116.

[21] B.M. Dawant, R. Li, B. Lennon, and S. Li, "Semi-automatic Segmentation of the Liver and its Evaluation on the MICCAI 2007 Grand Challenge Data Set," *Interface*, 2007, pp. 215-221.

[22] M. Freiman, L. Joskowicz, and J. Sosna, "A variational method for vessels segmentation: algorithm and application to liver vessels visualization," *Proceedings of SPIE*, 2009, pp. 72610H-72610H-8.

[23] S. Esneault and C. Lafon, "Liver Vessels Segmentation Using a Hybrid Geometrical Moments/Graph Cuts Method.," *IEEE transactions on bio-medical engineering*, vol. 33, 2009, pp. 1-19.

[24] M. Erdt, M. Raspe, and M. Suehling, "Automatic Hepatic Vessel Segmentation Using Graphics Hardware," *Lecture Notes in Computer Science*, vol. 5128, 2008, p. 403–412.

[25] D. Lesage, E.D. Angelini, I. Bloch, and G. Funka-Lea, "A review of 3D vessel lumen segmentation techniques: models, features and extraction schemes.," *Medical image analysis*, vol. 13, 2009, pp. 819-45.

[26] C. Kirbas and F. Quek, "Vessel extraction techniques and algorithms: a survey," *Third IEEE Symposium on Bioinformatics and Bioengineering, 2003. Proceedings.*, 2003, pp. 238-245.

[27] C. Kirbas and F. Quek, "A review of vessel extraction techniques and algorithms," *ACM Computing Surveys*, vol. vol, 2004, pp. 36no2pp81-121.

[28] R. WS, "ImageJ," *Bethesda, MD: US National Institutes of Health*, 1997.

[29] S. Pieper, M. Halle, and R. Kikinis, "3D Slicer," *IEEE International Symposium on Biomedical Imaging: Nano to Macro, 2004*, 2004, p. 632–635.

[30] K. Inc, "ParaView," *http://www.paraview.org*.

[31] P. Yushkevich, "ITK-SNAP," *www.itksnap.org*.

Information Piracy and Data incongruity prevention in WCMS to WCMS conversion procedure using precise information hiding and data integrity manager

Yeasin Ar Rahman
Department of Computer Science and Engineering
Chittagong University of Engineering and Technology
Chittagong, Bangladesh.
nibir201188@gmail.com

Khalid Ibn Zinnah
Department of Computer Science and Engineering
Chittagong University of Engineering and Technology
Chittagong, Bangladesh.
khalidex@yahoo.com

Abstract — This paper addresses the rapidly increasing issue of sensitive user data theft and the data incongruity due to Web Content Management System (WCMS) structure and very complex WCMS to WCMS conversion procedure. After retaining the full picture of the situation, the paper approaches a solution for securely accessing, modifying and migrating WCMS database(DB) using a precise information hiding and data integrity manager named Data Security Processor Software Unit (**DSPSU**) without compromising the sensitive data to fraudsters.

Index Terms — Encryption, End User Right, Database, Open-Source, WCMS, Web Security.

I. INTRODUCTION

Nowadays internet has become the most important organ of modern communication and information technology. It has enabled billions of people around the world to connect, share and communicate. As web publishing technology necessitates certain skills on HTML, CSS, PHP and working knowledge over server and networks, it is not very accessible to all kind of users. Thus users felt the need of automated tools to do all of these things easily. WCMS is server based software that has reduced this lengthy procedure of building, creating and publishing web content into a simple click [1]. Now majority of the users (or website administrators) use open source WCMS as it offers large user base, frequent updates and latest features. Most importantly it is free. Although open source WCMSs have evolved and the functionality has increased rapidly in these past few years, one of the major concern over open source WCMS is Security [2]. All of the source code of these WCMS can be found online, it is possible for a potential data thief to find a exploit in the source code to hack into the control panel. Even more problematic is that majority of the websites running these WCMS have no enforced privacy policy. So the security of sensitive user data is entirely up to the site owner [3].

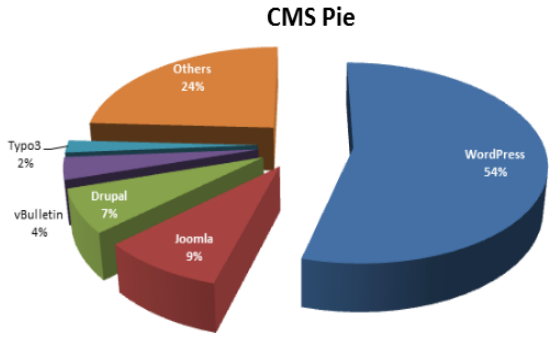


Figure 1: Percentage of users using WCMS [4,5]

It is possible for him to use this data for evil purpose. Another major aspect of WCMS security is that majority of the site hold raw data in the DB. And the username password associated with the WCMS DB is by default written inside the source code.

```
<?php
/*The base configurations of the WordPress.
 * This file has the following configurations: MySQL settings
 * Secret Keys, WordPress Language, and ABSPATH. You can find
 * by visiting {@link http://codex.wordpress.org/Editing_wp-c
 * wp-config.php} Codex page. You can get the MySQL settings
 * This file is used by the wp-config.php creation script dur
 * @package WordPress
 */
// ** MySQL settings - You can get this info from your web ho
/** The name of the database for WordPress */
define('DB_NAME', 'here_goes_database_name');
/** MySQL database username */
define('DB_USER', 'here_goes_username');
/** MySQL database password */
define('DB_PASSWORD', 'here_goes_password');
/** MySQL hostname */
define('DB_HOST', 'here_goes_host_ip');
/** Database Charset to use in creating database tables. */
define('DB_CHARSET', 'utf8');
/** The Database Collate type. Don't change this if in doubt.
define('DB_COLLATE', '');
?>
```

Figure 2: Source code for wp-config.php on Wordpress [6]

Open Source WCMS are maintained entirely by volunteers so the development and the update often stops due to the lack of user curiosity and drop in popularity.

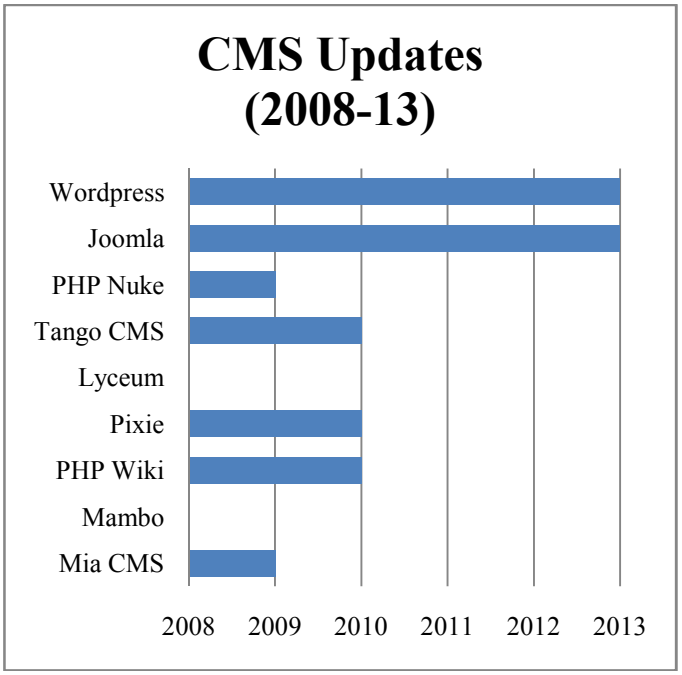


Figure 3: CMS update Rate till 2013[7]

Considering the reasons to use open source WCMS it is safe to assume that majority of the site owners do not have necessary knowledge to securely install a WCMS or secure a WCMS DB And they necessary do not have the knowledge to migrate to another WCMS [8]. For most of the open source WCMS, official support for upgrading or migrating to other WCMS is nearly nonexistent. Automated migration tools are not present in maximum cases and the process usually cannot transfer all data so data integrity is lost. The process is complex and so the owner (webmaster) definitely needs technician in the field to transfer data from one WCMS to another. The process is costly, time consuming and is not secured at all. The technical Specialist will get raw user data from the databases which might contain user sensitive or confidential information. Due to the cost of the process which increases exponentially according to number of users, many website owners decide to shut down the website and take large amount of searchable web content offline.

This paper focuses how sensitive data can be protected in an automated way in order to maintain the ease of use of the WMCS software. It further explains how the same technology will make sure the portability of the DB and so it will make sure that data is not corrupted or deleted.

This paper is organized as follows first it describes the possible way of implementing a software solution to address the problems. After that it evaluates the solution against the threats it will face. Then it describes the possible outcome of implementing the solution in a typical website. Finally a conclusion is given describing how necessary it is to implement a feasible data storing technique to ensure data security. Then possible future works and references are associated.

II. Procedure

All the WCMS maintain a hierarchical data management tables in the DB [6]; these tables hold all the necessary information of the WCMS. Different WCMS implement different structures to serve its purpose, so it not possible to directly migrate or convert the DB table. If an intruder gets direct or indirect access to these files then he might be able to deface, steal or drop all the content of the DB. Primary methodology for the DSPSU system is to encrypt all the data and associate it with an access key and save the encryption key in the cloud server. **DSPSU** will make the automated calls for sensitive data. The data will be encrypted with 128 bit AES [9] algorithm and will have different encryption keys which will be given by the users. The entire encryption key set will be stored in the cloud server and the cloud server will provide a access keys for the DB. Every time there will be a call for these data by the system which will go through the DSPSU. As the data gets routed in the cloud server, it is impossible for a potential data hijacker to steal information unless he knows the combination of all the encryption keys. The most important fact of the DSPSU is, it will never decrypt data from the DB unless the owner of the particular data specifically orders to do it.

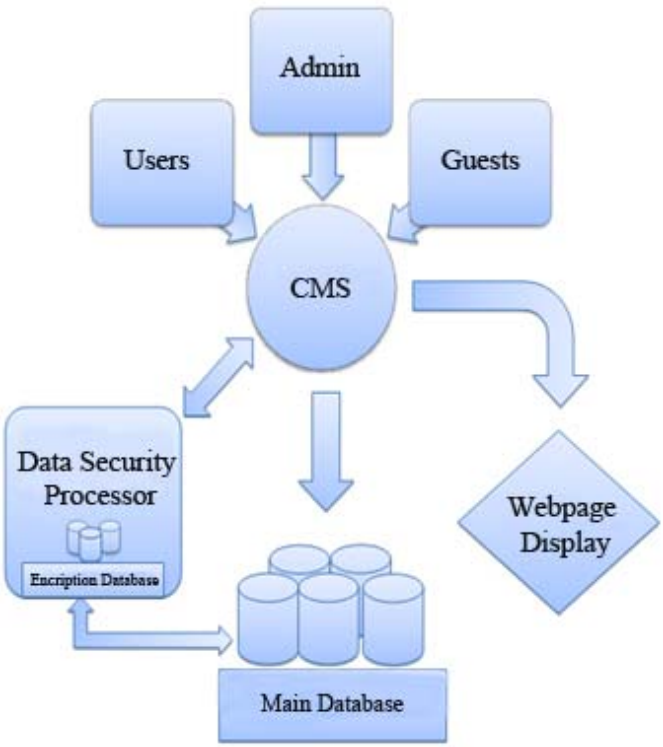


Figure 4: DSPSU Structure in WCMS

III. Working Principle of DSPSU

In order to maintain strong data integrity DSPSU maintains a set of working principle these are following:

- Identify User Name Password and associated user-data, post id, Post data, user post distinction or separation from the main database of the WCMS
- Encrypt all the sensitive data [email and user posts(if not public)]
- Sends the encryption keys to the cloud DB and accepts access key from the cloud server
- Associate access key with the main WCMS DB

After completing the encryption pass the DSPSU works like a bridge between the DB and the WCMS.As it creates interconnection between various DB tables it can also be used as a migration tool for the WCMS. The flowchart for the data migration procedure is given below

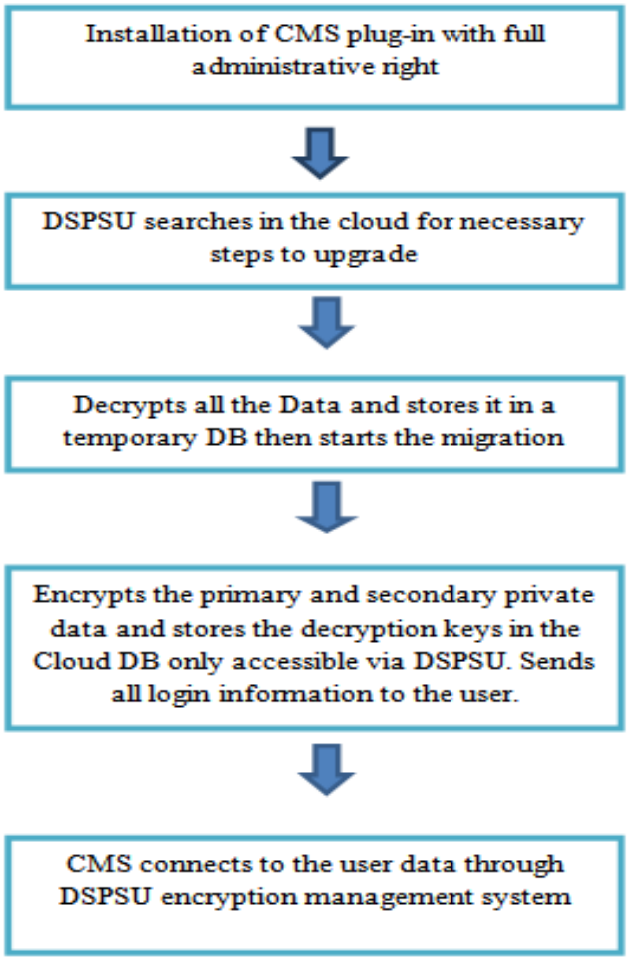


Figure 5: DSPSU Data Migration Flowchart

The necessary processes to migrate from a WCMS is called from the cloud server because the WCMS source code is often modified and sometimes the tables of the main DB changes, So to always maintain up to date information, it is necessary that the DSPSU procedure can be dynamically changed.

IV. Implementation of DSPSU in a WCMS

At the time of installation, DSPSU will create several new tables in the DB. A temporary group of tables will be created to store the temporary data at the time of primary installation. The data will be always backed up so it can retrieve if the results are not satisfactory. According to Rjandael algorithm (AES) [9], secure pass key generated by security processor which will be private and not accessible even by administrator from the WCMS control panel.

The encryption process will only occur to private or sensitive data, public post will remain unaffected.

In the specification of Rijndael, a byte with bits $b_7b_6b_5b_4b_3b_2b_1b_0$ maps to an element in $GF(2^8)$ given by the following polynomial:

$$b(x) = b_7x^7 + b_6x^6 + b_5x^5 + b_4x^4 + b_3x^3 + b_2x^2 + b_1x + b_0. \quad (2)$$

```
Round(State, ExpandedKey[i])
{
SubBytes(State);
ShiftRows(State);
MixColumns(State);
AddRoundKey(State,ExpandedKey[i]);
}
FinalRound(State,ExpandedKey[Nr])
{
SubBytes(State);
ShiftRows(State);
AddRoundKey(State,ExpandedKey[Nr]);
}
```

Figure 5: The Rijndael(AES) round transformation [9]

Finally, for the WCMS to WCMS migration step; DSPSU will generate common database. It will gather the necessary information from the cloud server and then complete the migration procedure. Then assign the value to NULL or default for which data does not exist from the existing WCMS. After finishing the migration the DSPSU will again encrypt the data for data security purpose.

V. Security Evaluation of DSPSU

Any sensitive data given by the user is automatically stored by the main WCMS DB [10]. So, an information hijacker will try to get access into this database to exploit or dump the data [11,12]. And a technical specialist who is appointed for migration the database into another WCMS will also get access to this DB.

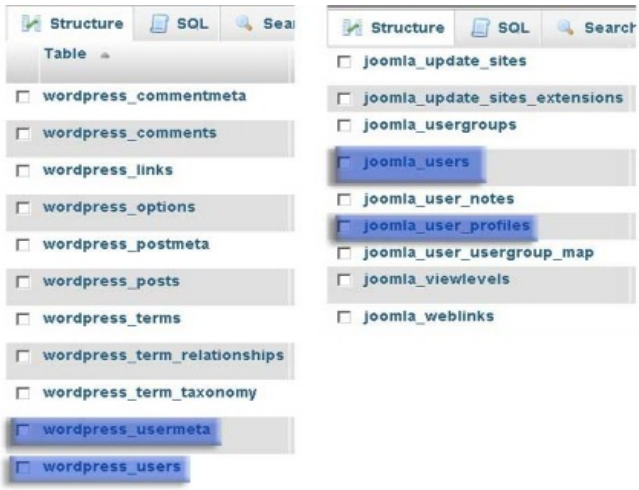


Figure 6: Database Structure of two popular WCMS (Wordpress and Joomla)[6]

AES encrypts data using keys given by the user, which does not exist in the DB by default. Here is a sample how DSPSU stores data in the DB.

Figure 7: User Data after DSPSU pass

As it can be seen, the sensitive data of the database is encrypted so a person cannot retrieve from it without knowing the password, thereby securing the data.

VI. OUTCOMES

The outcome of our study is that, after using DSPSU in a typical WCMS the user would not be needed to worry about Securing personal data. WCMS DB will be much more robust. Terabytes of valuable information can be saved and end users will face much less spam. However a possible negative impact of DSPSU is it needs access to at least one cloud server. Although these data can be stored in the default server, it can reduce its effectiveness greatly. Therefore further study is necessary to provide a more user friendly data storage.

VII. CONCLUSION

In this paper we have presented a simplified yet powerful approach toward WCMS security. The process we have mentioned is easily doable and is very cost effective. However, maintaining the usability of WCMS and at the same time securing data are two opposite forces [13,14]. Our study shows a potential model to solve the data security problem in WCMS. Nevertheless, from our study, we can conclude that more work is needed in order to develop a standard of storing data in WCMS software.

VIII. FUTURE WORK

As DSPSU seats on top of WCMS DB and it needs certain data verification from the cloud, a security certification can be authorized so that users can identify websites that respects data privacy, integrity and security. Further an organization can be established to maintain the certification process and also to continuously develop, upgrade and maintain the relevancy DSPSU.

IX. REFERENCE

[1] . Gonzalo Luzardo,Allan Avendaño,Katherine Chiluiza, Nervo Verdezoto. “A Web Content Management System to Create Accessible Web Sites” (ASSETS August 10, 2010).

[2] . Ganesh Vaidyanathan, Steven Mautone. “Security in dynamic web content management system applications”. Pages 121-125. ACM New York, NY, USA. Published 12, December 2009

[3] . Hua-Wei Mei, Tian Li, Ming-Quan Zhan. “Security Evolution Algorithm for enterprise class CMS Security system”. Proceedings of the Seventh International Conference on Machine Learning and Cybernetics, Kunming, 12-15 July 2008

[4] . W3techs web Technologies Surveys. Published on content management section (June 30, 2013)

[5] . Levia IT Technology Survey, Updated on September 19, 2012

[6] . Wordpress Source code. Updated June 21, 2013 Joomla Source code. Updated June, 2013

[7] . List of content management systems.(Updated May-June , 2012) (Wikipedia)

[8] . Meike, M, Sametinger, J., "Security in Open Source Web Content Management Systems". (Pages 44-51) published on August 04, 2009.

[9] . Henk C.A. van Tilborg. Encyclopedia of Cryptography and Security (Springer, 2005). Eindhoven University of Technology, Netherlands. Pages: 522-523.

[10] . OWASP. "The Ten Most Critical Web Application Security Vulnerabilities." OWASP Whitepaper, version 1.0, 2003.

[11] . Gonzalo Luzardo,Allan Avendaño,Katherine Chiluiza, Nervo Verdezoto. "A Web Content Management System to Create Accessible Web Sites" (ASSETS August 10, 2010).

[12] . Patel, S.K.; Rathod, V.R.; Prajapati, J.B., "Comparative analysis of web security in open source content management system," *Intelligent Systems and Signal Processing (ISSP), 2013 International Conference on* , vol., no., pp.344,349, 1-2 March 2013

[13] . Savan patel," Joomla, Drupal and WordPress- A statistical Comparison of Open Source CMS", Proceedings of 3rd international conference on Trendz in Information Sciences and Computing (TISC), 2011, IEEE conference publication, pp 182-187.

[14] . Meier, J.D., Mackman, A., Vasireddy, S. Dunner,M.,Escamilla, R., Murukan, A. "Improving Web Application Security—Threats and Countermeasures." Microsoft Corporation, 2003.

Linked Open Data Representation of Historical Heritage of Bangladesh

Shima Chakraborty
Dept. of Computer Science & Engineering
University of Chittagong
Chittagong - 4331, Bangladesh
Email: shima_csecu@yahoo.com

Md. Hasan Hafizur Rahman
Dept. of Computer Science & Engineering
University of Chittagong
Chittagong-4331, Bangladesh
Email: hasancse03@gmail.com

Md. Hanif Seddiqui
Dept. of Computer Science & Engineering
Bangabandhu Sheikh Mujibur Rahman Science and Technology University, Gopalganj-8100
and University of Chittagong, Chittagong-4331
Email: hanif@cu.ac.bd

***Abstract*—The historical heritage is a vast domain that generates distributed and heterogeneous data repositories in the existing World Wide Web (WWW). The integration of these heterogeneous data repositories and retrieving data from WWW using traditional search engine is a challenging task. In this regard, our research is to demonstrate the feasibility of semantic web technology for creating machine understandable description of entities associated with the creation and maintenance of historical heritage of Bangladesh. Moreover, our proposed approach uses the essence of semantic knowledge base composed of real historical heritage data that are interlinked with data from the $Geo-Bangladesh$, a machine understandable Linked Open Data (LOD) repository of geographic related data to the administrative structure of Bangladesh and retrieve specific information using semantic search tools like SPARQL. We perform a number of experiments with SPARQL to retrieve and inference specific information that shows usefulness of our proposed Linked Open Data on historical heritage of Bangladesh.**

I. Introduction

In the ongoing WWW, reduplication of heterogeneous content of historical heritage gradually increases redundant data on the web. These heterogeneous web contents are originating using HyperText Markup Language (HTML)[1] and visualizing to a user in the form of web pages. In this technology, we use predefined tags to define the style of sequence of characters in a web page rather than defining the meaning of these characters. These web content link one page to another page through the use of a hypertext link to create a chain of relevant and irrelevant pages in the WWW. However, these hypertext links in a page allow users to traverse information space using web browsers, while search engines index these pages and analyze the structure of links to infer potential relevance of users' search queries **[1]**. This role is empowered by the generic, accessible and extensible nature of the Web **[2]**, which is a key feature of the web's abandoned growth. Moreover, the Dynamic HTML (DHTML)[2] is a collection of technologies used to create interactive and animated web pages by using the combination of static markup language, a client side scripting language such as JavaScript [3], a presentation definition language such as Cascading Style Sheet (CSS)[4], and the Document Object Model (DOM)[5]. Although these technologies create web pages more interactive, however these are incompatible to define the semantic of data in a document. The representation of data using these technologies are easily human readable and interpret-able however, these are hardly understandable to machines. As a consequence of using these technologies, the traditional search engine retrieves millions of web page links for a user query rather than the more specific one. Furthermore, Extensible Markup Language (XML)[6] is a semi-structured data representation technique where users define their own tags to represent data in a hierarchical order. However, these user defined tags are lack of uniqueness to represent data for a domain of interest. Under these circumstances, machines can hardly understand the semantics of data, even for developing a smarter method to find more specific information on the web. Therefore we use structured database with these technologies in the back-end for a specific domain to originate the dynamic web so that users retrieve and update their domain information. Moreover, the database schema structures are restricted to define their attributes and their respective constraints in the database. The representation of these database schema are not adequate to develop meaningful associations among data in these repositories to create link with other data sources on the web. Therefrom humans or machines can understand these structured content effortlessly but machines are impotent to extract the meaning from these data representation on the web.

Therefore, information relevant to the historical heritage contains many specific characteristics in different formats such as multi-format, multi-topical, multi-lingual, multi-culture and multi-targeted **[3]**. Alongside of these multi-representational format of data on the web, a syntactic level problem is occurred for these contents when we define different character sets and data formats **[4]**. These different metadata formats are interpreted separately and data is encoded at different level of exactness. Furthermore, these domain specific web contents are managed centrally and these contents are spread over on the web. As a result of these distributed data sources on the web are insufficient to locate relevant data sources for retrieving particular information of a user's query using keyword based traditional search engine **[5]**. To solve these problem we define domain specific information with semantically heterogeneous

[1] http://www.w3.org/html
[2] http://en.wikipedia.org/wiki/Dynamic_HTML
[3] https://developer.mozilla.org/en-US/docs/Web/JavaScript
[4] http://www.w3.org/Style/CSS/Overview.en.html
[5] http://www.w3.org/DOM/
[6] http://www.w3.org/XML/

978-1-4799-3497-3/13/$31.00 ©2013 IEEE

and create highly interlinked data to change the scope of the current WWW. In this regard, enough exertion has been made to originate the next generation web, $Semantic\ Web$, is a term coined by Tim Berners-Lee et al **[6]**, define the semantic structure of data in a document so that these data are understandable to machines as well as to humans and make these data more visible **[7], [8]** to a search engine.

Furthermore, the collection of historical heritage information of Bangladesh is found as an environment for evaluating semantic web technologies. These technologies are used to extend the principles of the web from documents to data so that a machine can understand the meaning of these data **[9]**. In this regard, the main contribution of our research is to describe historical heritage entities using Resource Description Framework (RDF) **[10], [11]** and then description of these entities are added in a document as semantic markup using RDF in Attributes (RDFa)[7] to structure a website so that it can be understandable to machines to infer influential information. This semantic markup technique for the Extensible Hyper Text Markup Language (XHTML)[8] is used to create human-readable information with machine-understandable indicators for browsers and other programs. Later on we cross-examined our historical heritage information with one more data source **Geo-Bangladesh**, *a knowledge-base machine understandable data repository for administrative structure of Bangladesh* to retrieve location of all historical heritage of Bangladesh. Moreover, this process locate heritage in a map to increase search ability by mass people who have access to mobile technology. Therefore, we performed a number of experiments on our semantic database using SPARQL **[12], [13]** to retrieve and inference specific information to show the effectiveness of our Linked Open Data (LOD) **[14]** application.

The remainder of this paper is structured as follows. Section II introduces the general terminologies to perceive the consequent contents of this paper. The present structure of our domain is described in Section III, while Section IV focuses on the approach of our system. Therefore, Section V describes the procedure of making domain data in machine understandable format and the consecutive Section VI defines semantic data source, $Heritage - BD$ as Linked Open Data (LOD). Finally Section VII describes concluded remarks along with some future directions of our work.

II. General Terminologies

This section familiarizes some basic terminologies with their notion to readers those are used throughout this paper. It includes the theme of ontology, Resources, Resource Description Framework (RDF), Uniform Resource Identifier (URI), RDF in Attributes (RDFa) and Linked Data to comprehend the essence of machine understandable historical heritage information representation of Bangladesh.

A. Ontology

In the semantic web, an ontology is an explicit, formal specification of a shared conceptualization of a domain of interest **[15], [16]**. It represent knowledge in a formal way by defining identical meaning of concepts in different data sources on the web. It acts as the backbone of the semantic web vision **[17], [18]** which is considered as the next generation web. An ontology contains core ontology, logical mapping, knowledge base, and lexicon. Furthermore, a core ontology, S, is defined as five tuples

$$S = (C, \leq_C, R, \sigma, \leq_R) \tag{1}$$

Where C and R are two disjoint sets called "concepts" and "relations" respectively. A relation is known as a property of a concept. A function represented by $\sigma(r) =< dom(r), ran(r) >$ where $r \in R$, domain is dom (r) and range is ran (r). A partial order $\leq_R$ represents on R, called relation hierarchy, $r_1 \leq_R r_2$ iff $dom(r_1) \leq_C dom(r_2)$ and $ran(r_1) \leq_C ran(r_2)$. The notation $\leq_C$ represents a partial order on C, called concept hierarchy or "taxonomy". In a taxonomy, if $c_1 <_C c_2$ for $c_1, c_2 \in C$, then c_1 is a subconcept of c_2, and c_2 is a superconcept of c_1. If $c_1 \leq_C c_2$ and there is no $c_3 \in C$ with $c_1 <_C c_3 <_C c_2$, then c_1 is a direct subconcept of c_2, and c_2 is a direct superconcept of c_1 denoted by $c_1 < c_2$ **[19]**.

B. Resources

Things, *items of interest of a domain* published on the web containing properties and relationships are known as resources. Resources are classified as information and non-information resources. Information resources are documents, images and other media files and on the other hand, non-information resources referred to all real world objects exist outside of the web such as places. In the semantic web technology, we use Hyper Text Transfer Protocol - Uniform Resource Identifiers (HTTP-URIs) for every resources in RDF to identify on the web of data.

C. RDF Data Model

The Resource Description Framework (RDF) is a data model that expressed knowledge pieces in the subject-predicate-object form known as triples **[20]**. The Fig. 1 demonstrates the basic structure of triple for the knowledge *"Shahid Minar is a Monument"* to comprehend the semantic knowledge representation in Linked Data.

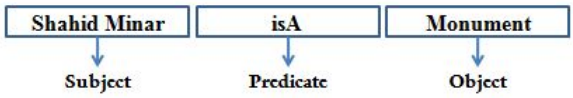


Fig. 1. RDF Triple for *"Shahid Minar is a Monument"*

Furthermore, this model is a generic, graph-based framework to structure and link data that describe resources of the world **[2]**. The graph representation is depicted in the Fig. **??** for the triple given in the Fig. 1.

D. RDF in Attributes (RDFa)

Resource Description Framework in Attributes (RDFa) is used to add semantic information to the XHTML markup to enhance the visual information on the web. The Fig. 2 demonstrates the use of semantic markup using RDFa to comprehend machine understandable representation of data in a document.

[7] http://www.w3.org/TR/xhtml-rdfa-primer/
[8] http://www.w3.org/TR/xhtml1/

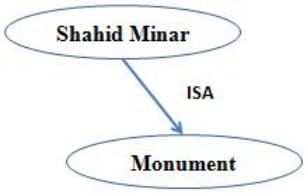


Fig. 2. RDF Graph Representation for the Triple in Fig. 1.

In this representation there are two articulation, one is for traditional visualization of data however, another is for semantic markup using RDFa representation. The traditional visualization represent in a web page $Type : Historical\ Heritage$ and $Name : Central\ Shahid\ Minar$ creating two paragraph. This representation is easily understandable to humans, but hardly understandable to machines. Moreover, the other visualization using RDFa represent similar content those are easily understandable to machines and humans.

<html> <body> ***Type: Historical Heritage*** ***Name:Central Shahid Minar*** </body> </html>	<html> <body> <h2 property = **"http://purl.org/dc/elements/1.1/type"**>*Historical Heritage*</h2> Type:<span property = **"http://purl.org/dc/elements/1.1/title"**>*Central Shahid Minar* </body> </html>
Traditional HTML Tags	**RDFa Semantic Markup**

Fig. 3. Data Representation using Traditional and RDFa Technology

E. Linked Data

The ontology knowledge base is a structure

$$KB = (C, R, I, \imath_C, \imath_R)$$

consisting of two disjoint sets C and R as defined before, a set I whose elements are called instances, two functions $\imath_C$ and $\imath_R$ are called concept instantiation and relation instantiation respectively.

With the structure of ontology and ontology knowledge base, we publish explicitly defined data on the web so that it is understandable to machines and these data create link to data on diverse knowledge repositories on the web **[2]**. Under this circumstances information representation as Linked Data spread the web with a global data space connecting data from various data sources. These data sources rely on documents that contain data in RDF, those are serialized in different formats such as N-TRIPLE[9], RDF/XML **[21]** and RDFa **[15]**. The representation of these data in RDF create typed links between arbitrary related resources in the web **[2], [14]**. Therefore, linked data relies on two technology to locate useful information from the diverse knowledge-base data repositories. These are URI (Uniform Resource Identifiers) and HTTP (Hyper Text Transfer Protocol) **[22]**. HTTP-URIs make good names for two reasons: they provide a simple way to create globally unique name without centralized management; and URIs work not just as a name however, as a means of accessing information about a resource over the web.

[9]http://www.w3.org/TR/n-triples/

III. Present Structure of the System

Bangladesh is a nature of beauty with its various culture and heritage. All historical heritages of Bangladesh are spread over on the geographical coordinates from $20^0 34N$ to $26^0 38N$ latitude and from $88^0 01E$ to $92^0 41E$ longitude. Moreover, the area of this country is organized in an administrative structure from upper level administration, $divisions$ to lower level administration, $villages$. In this structure, there are seven first level administrative divisions farther redivided into districts. Then sixty four (64) districts are sub-divided into upazila or thana. Except police stations in metropolitan area are partition into several unions and each union consisting of villages those are attached to each mouza. In the metropolitan area, each police station is divided into wards, those are sub-divided into mohollas. In the third level administrative structure, there are more than 500 upazilas and 509 administrative thanas for a total of more than 1009 sub-districts. Finally there are more than eighty seven thousand (>87,000) instances of $village$ in Bangladesh. The instances of village and moholla are populated place consist all historical heritage sites throughout the country. The logical relationship among different levels of the administrative structure of Bangladesh is portrait in the Fig. 3. In spite of these logical relationship it also demonstrates the association of historical heritage sites with administrative structure of Bangladesh.

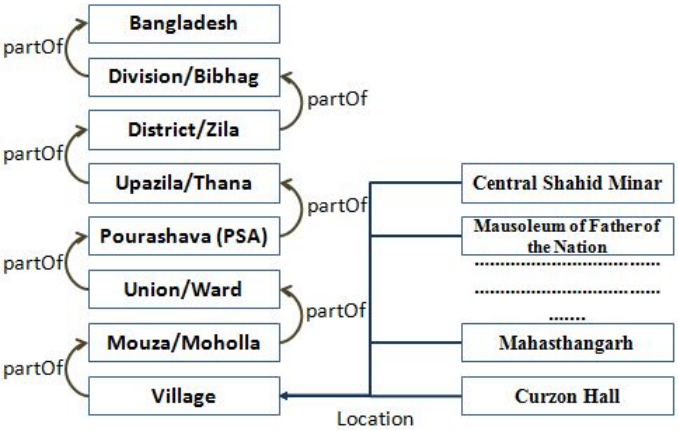
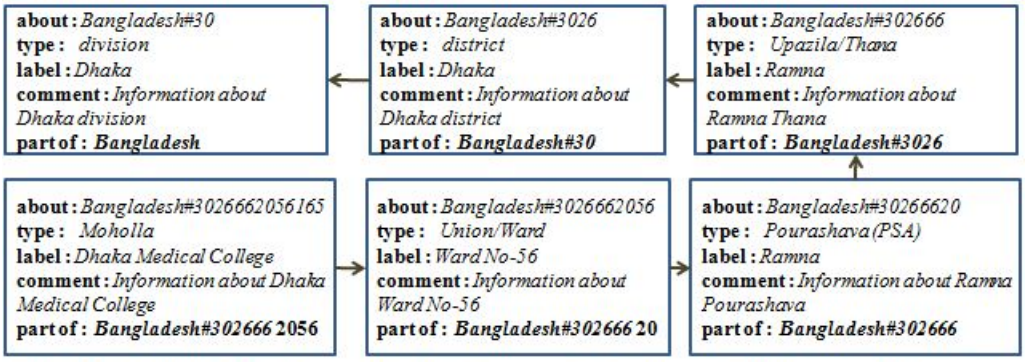


Fig. 4. Administrative Structure of Bangladesh with Logical Relationship

Moreover, the Fig. 4 manifest the link of moholla ($Dhaka\ Medical\ College$) to division ($Dhaka$) which demonstrates data to data link of the administrative structure of Bangladesh. However, the historical Heritage $Central\ Shahid\ Minar$ is attach to the area $Dhaka\ Medical\ College$ and the similar adhere for every heritage of Bangladesh.

Therefore, there are various heritage related to historical and religious places, historical monuments and archaeological sites spread all over Bangladesh. The hierarchical arrangement of heritage structure is outlined in the Fig. 5. Moreover, desolate ruins are still visible in many places due to many illustrious dynasties of kings and sultans ruled and disappeared and left their mark in the shape of magnificent cities and



Bangladesh	Dhaka Medical College	Ward No-56	Ramna Pourashava	Ramna Thana	Dhaka District	Dhaka Division
Geo-code	165	56	20	66	26	30

Fig. 5. Data to Data Link between Administrative Structures of Bangladesh

monuments. In these case new information and fresh evidence of heritage are gradually added on the web. These fresh explorations are added substantially to our knowledge of the history and chronology of ancient Bangladesh and various aspects of her life and culture. These heritage information are described using traditional technologies on the web. The representation of data using these traditional technologies are not machine understandable due to the lack of sufficient explicit meaningful description of data. In these case machines can not extract meaning of these data to increase information visibility on the web. In this regard, a semantically rich historical heritage data repository is essential to achieve the interoperability on the web among compatible knowledge-base data sources using common vocabularies to reduce the procreation of redundant data on the web.

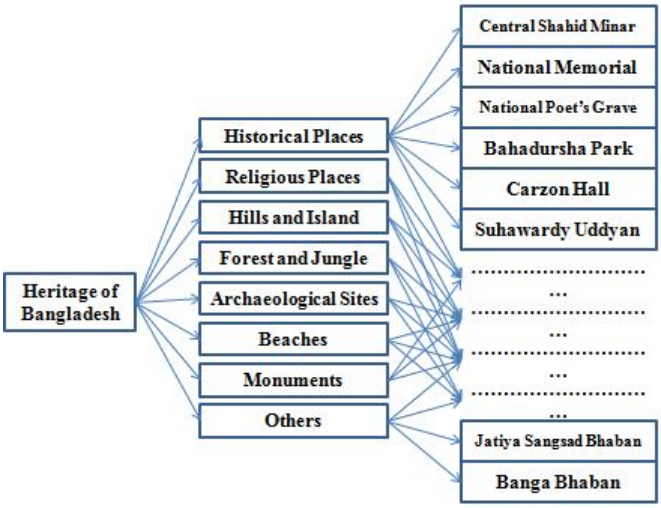


Fig. 6. Hierarchical Structure of Heritage in Bangladesh

IV. Our Approach

The approach of our system is to develop an application using semantic technology for historical heritage to retrieve and inference more specific information in favor of a users query. In this regard, we identify the apposite resources and then define concepts and properties of these resources and their relationship to build a knowledge repository. Therefore, we describe each resource using RDF technology and originate these description in a document using RDFa technology so that these representation can understand by machines as well as to humans. Furthermore, we published our historical heritage knowledge repository $Heritage - BD$ as LOD so that we share resources with other compatible data sources to address the interoperability issue in an efficient way. Therefore, we cross-examined our web contents with one more knowledge data repository $Geo - Bangladesh$, to retrieve the geographic location of a historical heritage. In the next step, we perform a number of experiments with SPARQL[10] to retrieve and inference more specific information on the web. The Fig. 6 demonstrates the overview of our approach to understand the significance of the system.

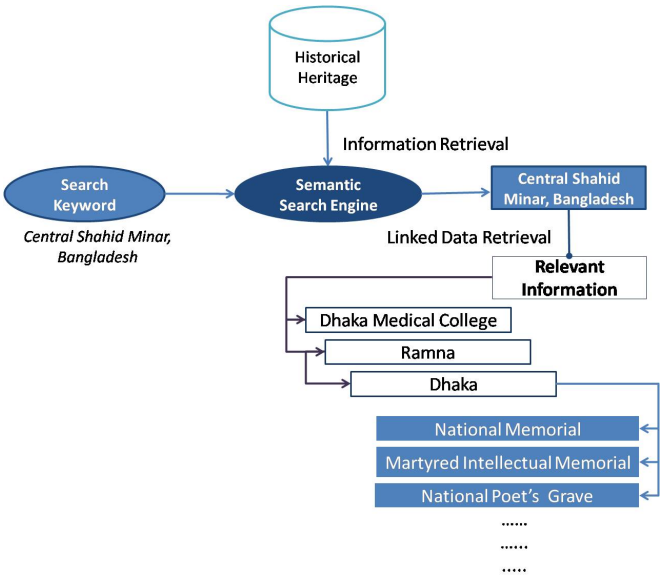


Fig. 7. Overview of our System

[10]http://www.w3.org/TR/rdf-sparql-query

Subject	Predicate	Object
<http://www.skeim.org#30>	<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>	<http://www.skeim.org#division>.
<http://www.skeim.org#30>	<http://www.w3.org/2000/01/rdf-schema#label>	"Dhaka @en".
<http://www.skeim.org#30>	<http://www.w3.org/2000/01/rdf-schema#comment>	"Information about Dhaka division @en".
<http://www.skeim.org#30>	<http://www.w3.org/2003/01/geo/wgs84_pos#lat>	"24.16667@en".
<http://www.skeim.org#30>	<http://www.w3.org/2003/01/geo/wgs84_pos#long>	"90.41667@en".
<http://www.skeim.org#30>	<http://www.w3.org/2000/01/rdf-schema#partOf>	<http://www.skeim.org#Bangladesh>.
<http://www.skeim.org#3026>	<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>	<http://www.skeim.org#district>.
<http://www.skeim.org#3026>	<http://www.w3.org/2000/01/rdf-schema#label>	"Dhaka@en".
<http://www.skeim.org#3026>	<http://www.w3.org/2000/01/rdf-schema#comment>	"Information about Dhaka district @en".
<http://www.skeim.org#3026>	<http://www.w3.org/2003/01/geo/wgs84_pos#lat>	"23.81092@en".
<http://www.skeim.org#3026>	<http://www.w3.org/2003/01/geo/wgs84_pos#long>	"90.36542@en".
<http://www.skeim.org#3026>	<http://www.w3.org/2000/01/rdf-schema#partOf>	<http://www.skeim.org#30>.

Fig. 8. Machine Understandable Representation of Data in N-TRIPLE format for the *Administrative Structure of Bangladesh*

V. Making Data Machine Understandable

Information representation using semantic web technology is machine interpretable **[9]**. This technology contains both metadata, *which is data about data,* associated with the web information and ontology of our domain. We describe each resources using common vocabularies those are wider consensus on the web according to the definition of our ontology. In this regard, we developed our domain ontology $Heritage - BDOnto$ using the ontology development environment $Protege$ [11] by defining their machine understandable definition of concepts with their properties and establish their relationship types among them. In spite of the ontology and resource description, we can originate these RDF description of each resources using RDFa in a XHTML document to increase the web sites visibility to a search engine as well as to share resources with other compatible knowledge-base data repositories on the web. Furthermore, the semantic markup for data in a document are described with the help of vocabularies on the web and understandable these data to humans as well as to machines **[23], [7]**. We use vocabularies from Dublin Core (DC)[12] and Friend-of-a-Friend (FOAF)[13] standard to encode meta-data of our domain. For example, a resource type is defined in a class using $http://www.w3.org/1999/02/22-rdf-syntax-ns\#type$ vocabulary. Each instance of a class consist their related features such as name and comment is described using $http://www.w3.org/2000/01/rdf-schema\#label$ and $http://www.w3.org/2000/01/rdf-schema\#comment$ vocabularies. Moreover, latitude and longitude of each location are defined using common vocabularies of $http://www.w3.org/2003/01/geo/wgs84_pos\#lat$ and $http://www.w3.org/2003/01/geo/wgs84_pos\#long$. The Fig. 7 gives a close look of the machine understandable data representation of the administrative structure of Bangladesh. Moreover, the representation of historical heritage information using RDF format is also demonstrates in the Fig. 8. The RDF representation of these data used as semantic markup in a document of RDFa description for *"Historical Heritage of Bangladesh"* is depicted in the Fig. 9 .

```
<html xmlns="http://www.w3.org/1999/xhtml"
    xmlns:foaf="http://xmlns.com/foaf/0.1/"
    xmlns:dc="http://purl.org/dc/term/"
    xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#">

<head>
<meta name="dc:title" content="Linked Open Data Representation of Historical Heritage of Bangladesh"/>
<meta name="dc:creator" content="Shima Chakraborty, Md. Hasan Hafizur Rahman, Md. Hanif Seddiqui"/>
<title property="dc:title">Historical Heritage of Bangladesh</title>
</head>

<body>

<rdf:Description rdf:about="http://www.skeim.org/Heritage-BD#302666205616501">
    <h2 property="rdfs:label">Central Shahid Minar@en </h2>
   <p><b>Abstract:</b> <dc:abstract>The Central Shaheed Minar is a national
          monument in Dhaka, Bangladesh, established to commemorate those
          killed during the Language Movement demonstrations of 1952. The day
          is also now observed as International Mother Language Day across the
          world. Hundreds and thousands of barefooted people with floral wreaths
          and bouquets gather at this monument from the first hour of 21st February
          every year to pay homage to the martyrs.</dc:abstract></p>

   <p><b>Type:</b> <a property =rdf:type href="http://www.skeim.org#
Monuments">Monument, Hostorical Heritage</a></p>

   <p><b>Location:</b> <a property=dc:location href=http://www.skeim.org/
Geo-BD#3026662056165>Dhaka Medical College Area</a></p>

   <p><b>creator:</b> <a property=dc:creator href="http://www.skeim.org/
Heritage-BD/Creator#S0101">Hamidur Rahman</a></p>

   <p><b>Inauguration Day:</b> <dc:date>February 21, 1963</dc:date></p>
 ..............................................................................
 ..............................................................................
 ..............................................................................
  </rdf:Description>
  ..............................................................................
 ..............................................................................
 ..............................................................................
 ..............................................................................
 ..............................................................................
 ..............................................................................
</body>
</html>
```

Fig. 10. Semantic Markup using RDFa Technology for *"Historical Heritage of Bangladesh"*

[11] http://protege.stanford.edu/
[12] dublincore.org/documents/dces/
[13] http://xmlns.com/foaf/spec/

Subject	Predicate	Object
<http://www.skeim.org/Heritage-BD#302666205616501>	<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>	<http://www.skeim.org/Heritage-BD#monument>.
<http://www.skeim.org/Heritage-BD#302666205616501>	<http://www.w3.org/2000/01/rdf-schema#label>	"Central Shahid Minar @en".
<http://www.skeim.org/Heritage-BD#302666205616501>	<http://www.w3.org/2000/01/rdf-schema#comment>	"Information about Central Shahid Minar of Bangladesh @en".
<http://www.skeim.org/Heritage-BD#302666205616501>	<http://purl.org/dc/ terms/abstract>	"The Central Shahid Minar is a national monument in Dhaka, Bangladesh, established to commemorate those killed during the language movement demonstrations of 1952. The day is also now observed as International Mother language day across the World. Hundreds and thousands of barefooted people with floral wreaths and bouquets gather at this monument from the first hour of 21st February every year to pay homage to the martyrs ".
http://www.skeim.org/Heritage-BD#302666205616501	<http://purl.org/dc/ terms/creator>	"Hamidur Rahman@en".
http://www.skeim.org/Heritage-BD#302666205616501	<http://purl.org/dc/ terms/Location>	<http://www.skeim.org/3026662056165>.

Fig. 9. Semantic Description for *"Historical Heritage of Bangladesh"* using RDF Framework

VI. Linked Open Data in Historical Heritage of Bangladesh

The semantic database $Heritage - BD$ share resources with other knowledge repository to create a chain of data on the web. Chaining different web content increases the number of knowledge-intensive task that can be carried out automatically [24]. The semantic web supports technology for describing, discovering and accessing web contents that formed a chain of interoperable web applications dynamically [25]. In this regard, our $Heritage - BD$ repository published as LOD by the outlines of Tim Berners Lee **[2]**. These outlines are - for each resources we use a URI and to look up these URI we need to follow the de-referenceable URIs (HTTP-URIs). Therefore, when we look up a specific URI then we get useful information and links to other resources to retrieve compatible information from the web of data.

Furthermore, our $Heritage - BD$ reuse resources $Location$ from $Geo - Bangladesh$ to reduce redundant information on the web and at the same time interoperability issues can be achieved. However, the information from the $Geo - Bangladesh$ consist latitude and longitude values to locate heritage in a map. The Fig. 10 depicts resource sharing approach from $Geo - Bangladesh$ with $Heritage - BD$.

Moreover, to achieve reasoning from knowledge-base data source, $Heritage - BD$, we use java based semantic web application development framework, $Jena$ [14] for rule based inference and $SPARQL$, a semantic search engine for RDF data-sets. In this semantic environment we retrieve specific information of a users query about historical heritages of Bangladesh from diverse LOD repositories on the web. The SPARQL query in the Fig. 11 is used to retrieve URI of Central Shahid Minar of Bangladesh and the query result is also demonstrates in the Fig. 12. Moreover, we performed experiments to retrieve all identical historical heritage sites spread over the country using inference engine and share resources with other compatible information sources successfully.

[14]http://jena.apache.org/

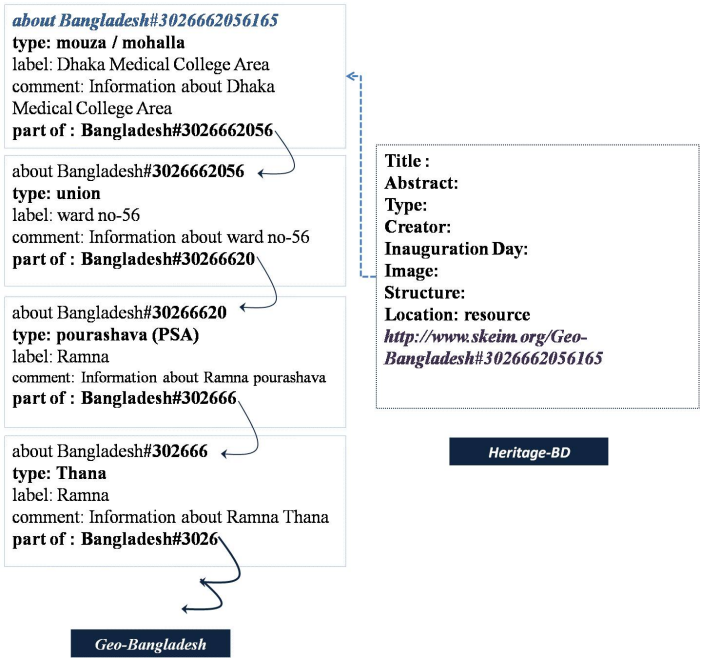


Fig. 11. The Linked Open Data (LOD) and resource sharing approach with $Geo - Bangladesh$ and $Heritage - BD$

```
PREFIX rdf: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX rdfs: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX Heritage: < http://www.skeim.org/Heritage-BD>
SELECT  ?URI
  WHERE
    {
        ?URI  rdfs: label "Central Shahid Minar@en".
        };
```

Fig. 12. Semantic query using SPARQL to retrieve URI of Central Shahid Minar.

URI
http://www.skeim.org/302666205616501

Fig. 13. Search result for the query in Fig. 11.

Furthermore, the snippet of machine understandable visual representation of historical heritage information of Bangladesh is portraying in the Fig. 13.

Central Shahid Minar

Abstract: The Central Shahid Minar is a national monument in Dhaka, Bangladesh, established to commemorate those killed during the Language Movement demonstrations of 1952. The day is also now observed as International Mother Language Day across the world. Hundreds and thousands of barefooted people with floral wreaths and bouquets gather at this monument from the first hour of 21st February every year to pay homage to the martyrs.

Type: Monument, Historical Heritage

Location: Dhaka Medical College Area

Creator: Hamidur Rahman

Inauguration Day: February 21, 1963

Image:

Structure: Shaheed Minar National Monument is 14 foot high stage which is constructed from columns of pure marble stone, with stairs and railings painted white and the fences on either side bearing lines of poetry penned by famous poets made out of iron letters.

Fig. 14. Visualized Representation of Historical Heritage Information of Bangladesh

VII. Conclusion

Machine understandable historical heritage information of a country is essential to spread its history and culture all over the web those attract visitors as well as interested people who are willing to explore their knowledge on them. The knowledge-base data repository of historical heritage decrease data profusion for other applications those use these compatible data. In spite of reduction of identical data on the web our semantic data repository $Heritage - BD$ retrieve geographical location of historical heritage site from one more machine understandable LOD repository $Geo - Bangladesh$ to address interoperability issue in an efficient way. Moreover, the semantic markup in our application increases the web visibility to a search engine. However, our proposed LOD application locate heritage in a map to increase the search ability by mass people who have access to mobile technology. In future we develop an intelligent agent to automatically crawl compatible data to augment heritage information and link with other knowledge base on the web.

References

[1] S. Brin and L. Page, "The anatomy of a large-scale hypertextual web search engine," *Computer networks and ISDN systems*, vol. 30, no. 1, pp. 107–117, 1998.

[2] C. Bizer, T. Heath, and T. Berners-Lee, "Linked data-the story so far," *International Journal on Semantic Web and Information Systems (IJSWIS)*, vol. 5, no. 3, pp. 1–22, 2009.

[3] J. Hendler and Y. Ding, "Synthesis lectures on semantic web: Theory and technology," 2012.

[4] P. Bertone, Y. Kluger, N. Lan, D. Zheng, D. Christendat, A. Yee, A. M. Edwards, C. H. Arrowsmith, G. T. Montelione, and M. Gerstein, "Spine: an integrated tracking database and data mining approach for identifying feasible targets in high-throughput structural proteomics," *Nucleic acids research*, vol. 29, no. 13, pp. 2884–2898, 2001.

[5] L. Galanis, Y. Wang, S. R. Jeffery, and D. J. DeWitt, "Locating data sources in large distributed systems," in *Proceedings of the 29th international conference on Very large data bases-Volume 29*. VLDB Endowment, 2003, pp. 874–885.

[6] T. Berners-Lee, J. Hendler, O. Lassila *et al.*, "The semantic web," *Scientific american*, vol. 284, no. 5, pp. 28–37, 2001.

[7] L. Yu, *A Developers Guide to the Semantic Web*. Springer, 2011.

[8] A. Graf, "Rdfa vs. microformats," *Digital Enterprise Research Institute, Innsbruck*, 2007.

[9] T. Berners-Lee, "Semantic web road map," 1998.

[10] I. Davis, "An introduction to rdf."

[11] B. McBride, D. Boothby, and C. Dollin, "An introduction to rdf and the jena rdf api," *Retrieved August*, vol. 1, p. 2007, 2004.

[12] L. Dodds, "Introducing sparql: Querying the semantic web," *Retrieved on April*, vol. 20, p. 2006, 2005.

[13] P. McCarthy, "Search rdf data with sparql: Sparql and the jena toolkit open up the semantic web, in developerworks," 2005.

[14] T. Heath and C. Bizer, "Linked data: Evolving the web into a global data space," *Synthesis Lectures on the Semantic Web: Theory and Technology*, vol. 1, no. 1, pp. 1–136, 2011.

[15] W. W. W. Consortium *et al.*, "Rdfa primer: Bridging the human and data webs," *Retrieved October*, vol. 4, p. 2009, 2008.

[16] R. Studer, V. R. Benjamins, and D. Fensel, "Knowledge engineering: principles and methods," *Data & knowledge engineering*, vol. 25, no. 1, pp. 161–197, 1998.

[17] T. A. Powell, "The complete reference html and xhtml," 2003.

[18] W. W. W. Consortium *et al.*, "Xhtml 1.0 the extensible hypertext markup language (2002)," *URL: http://www. w3. org/TR/xhtml1*.

[19] M. Ehrig, "Ontology alignment: Bridging the semantic gap, volume 4 of semantic web and beyond computing for human experience," 2007.

[20] G. Antoniou and F. Van Harmelen, *A semantic web primer*. MIT press, 2004.

[21] D. Beckett and B. McBride, "Rdf/xml syntax specification (revised)," *W3C recommendation*, vol. 10, 2004.

[22] T. Berners-Lee, "Linked data-design issues (2006)," *URL http://www. w3. org/DesignIssues/LinkedData. html*, 2011.

[23] T. R. Gruber *et al.*, "Toward principles for the design of ontologies used for knowledge sharing," *International journal of human computer studies*, vol. 43, no. 5, pp. 907–928, 1995.

[24] Y. Ding, D. Fensel, M. Klein, and B. Omelayenko, "The semantic web: yet another hip?" *Data & Knowledge Engineering*, vol. 41, no. 2, pp. 205–227, 2002.

[25] M. Daconta, L. Obrst, and K. Smith, "The semantic web: a guide to the future of xml, web services, and knowledge management," 2003.

Numerical Study of a Loaded U-Antenna for 3.5 GHz Mobile WiMAX and 7.5 GHz Military Satellite Communication Applications

Md. Juwel Rana
Dept. of Electrical & Electronic Engineering
Daffodil International University
Dhaka, Bangladesh
jreee07@gmail.com

Md. Abu Shahab Mollah
Dept. of Electrical & Electronic Engineering
Ahsanullah University of Science & Technology
Dhaka, Bangladesh
asm.eee@aust.edu

***Abstract*— This paper presents the numerical simulation of a low-profile dual frequency loaded U-antenna for 3.5 GHz WiMAX and 7.5 GHz military satellite communication applications. The antenna has a very compact size of 23.5×14 mm^2 and provides a bandwidth of 340 MHz (3.4 GHz-3.74 GHz) and 570 MHz (7.23 GHz-7.8 GHz) in the two different resonant modes. The antenna provides maximum gain of 7.51 and 3.14 dBi respectively at S_{11} < -10 dB. It also provides almost omnidirectional radiation characteristics and excellent impedance matching in the operating bandwidth.**

***Index Terms*— Loaded-U antenna, WiMAX, Numerical Electromagnetic Code (NEC)**

I. Introduction

Today, the blessings of wireless communication has rapidly widespread to the every inches of a human life. The world has experienced tremendous advancement and development in wireless communication service in the last decade and still continued for more. The thirst for new wireless capacity through deployment of greater intelligence in wireless network seems never to come to end. The future development of the personal communication devices will aim to provide image, speech and data communications at any time, and anywhere around the world [1]. Among the various applications of wireless communication, personal area network (PAN), wireless local area network (WLAN), worldwide interoperability for microwave access (WiMAX), other wireless local loop (WLL) and a variety of satellite systems are most common in modern technology. The rapid growing WLAN operating bands are IEEE 802.11 b/a/g at 2.4 GHz (2400–2484 MHz), 5.2 GHz (5150–5350 MHz) and 5.8 GHz (5725–5825 MHz) also the bands of WiMAX operation in the 2.5 GHz band (2500–2690 MHz), 3.5 GHz band (3300–3700 MHz) and 5.5 GHz band (5250–5850 MHz). In recent times WiMAX can provide a long operating range with a high data rate for mobile broadband wireless access, faultless internet access for wireless users becomes more popular [2]. Also, the military satellite communication uses the frequency in the X-band (7.9-8.4GHz for uplink and 7.25-7.75 GHz for downlink) assigned by the International Telecommunication Union. To utilize the wide range of applications of wireless communication efficiently urges on the need of dual band or multiband antennas. Recently, the design of dual band or multiband antennas has received the attention of antenna researchers. A Printed Multiband Antenna operating at 2.4/3.5/5.2/5.5/5.8 GHz for Cellphone Applications [3], Capacitively Fed Hybrid Monopole/Slot Chip Antenna for 2.5/3.5/5.5 GHz WiMAX Operation in the Mobile Phone [4], Dual Wideband Printed Monopole Antenna for WLAN/WiMAX Applications [5], Internal Composite Monopole Antenna for 2500–2690/3400– 3600/5250–5850 MHz WLAN/WiMAX Operation in a Laptop Computer [6], Multiband CPW-Fed Triangle-Shaped Monopole antenna for 2.4/3.4/5 GHz Wireless Applications [7] have been reported. As the need for more compact, reliable and multipurpose antenna is growing day by day, we concentrate our attention on simple wire antenna as it has some special features like small size, low cost and high gain properties.

In this paper, the simulation of a low profile, dual band and high forward gain performance for providing 3.5 GHz mobile WiMAX and 7.25 GHz military satellite communication applications is presented. The simulation is carried out in Numerical Electromagnetics Code (NEC).

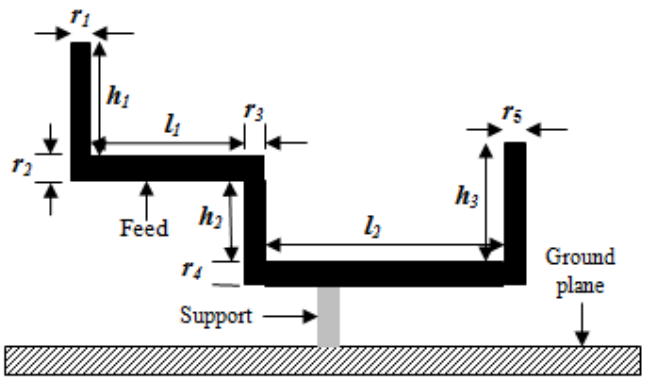


Figure 1. Loaded U-antenna

II. ANTENNA DESIGN

Taking a simple U-shaped antenna as a basic design desired for dual band operation which may cover the WiMAX and military satellite communication applications, all the possible modifications have been experimented. This antenna is studied in a trial and error basis to be ascertain of the effect of different loading on the antenna performance to find out the optimal design. Finally, the loaded U-shaped structure meets the specifications of dual band that provides the required applications. Fig. 1 shows the geometry of the loaded U-antenna. It mainly consists of a U-shaped structure and an L-shaped structure is placed on top of it as a load. The structure has a dimension of 23.5×14 mm^2. For this study, copper conductor is assumed and the antenna was intended to be matched to 50 Ω system impedance. The antenna is to be feed by 50 Ω coaxial connector, with its central conductor connected to the feeding point and its outer conductor soldered to the ground plane. The dimension of the ground plane considered as 60 mm × 60 mm. The lengths and diameters of the antenna as given in TABLE I.

TABLE I: DIMENSION OF PROPOSED ANTENNA

Parameter	Length (mm)	Dimension (mm^2)
l_1	10	23.5×14
l_2	13.5	
h_1	8.5	
h_2	5.5	
h_3	6.5	
r_1	0.61	
r_2	1.15	
r_3	0.56	
r_4	0.9	
r_5	0.48	

III. SIMULATION RESULT

The proposed antenna is constructed and numerically analyzed using NEC simulator. Fig. 2 shows the simulated frequency responses of return loss for the proposed design. The antenna provides appreciable return loss in the respective frequency bands than the commonly required level of S_{11} < -10 dB. It is seen that the antenna offers two resonant modes at 3.55 GHz and 7.5 GHz covering a bandwidth of 340 MHz (3.4-3.74 GHz) and 570 MHz (7.23-7.8 GHz). The peak values of return loss is -37.899 and -62.804 dB at lower and upper frequency bands respectively. To be certain of the optimal design, the antenna parameters like lengths, heights, radios etc. are varied and its performance is analyzed in NEC. Fig. 3 depicts this situation and shows that the proposed antenna parameters suits the optimal design.

Fig. 4 shows the variation of voltage standing wave ratio (VSWR) as a function of frequency. The VSWR was obtained as 1.026 and 1.001 at 3.55 and 7.5 GHz respectively. Simulation result shows that the value of VSWR is almost 1 and its variation within the frequency bands is very small. Fig. 5 shows the variation of input impedance with respect to frequency. From simulation result we found that the input impedance is 50.5 and 50.04 Ω at the respective bands, which is very close to 50 Ω. Thus the antenna can be well matched with the 50 Ω coaxial connector line. Fig. 6 presents the phase variation of the proposed antenna with respect to frequency. Simulation result shows that variation in phase is very small in all over the antenna bandwidth in both of the bands. Fig. 7 shows the variation in antenna gain of loaded U-shaped structure. It is found that the peak gain is obtained as 7.51 and 3.14 dBi at 3.55 and 7.5 GHz respectively. Gain variation is less than 0.1 and 1.17 dBi respectively in the operating bandwidth which indicates its stable operation for the mentioned applications. Peak gain comparison with the proposed antenna and reference antenna for mobile WiMAX application is listed in TABLE II. From the comparison table, it is evident that the proposed antenna has much higher gain than the antenna have been referred for mobile WiMAX operation. Fig. 8 Shows the normalized radiation patterns for the resonant frequency are shown as total gain in vertical (YZ/XZ plane) and horizontal plane (XY plane). The antenna's normalized total radiation in vertical and horizontal plane is almost omnidirectional at the 3.5 GHz WiMAX and 7.5 GHz military satellite communication operating frequency. The proposed antennas has radiation efficiency of 100%.

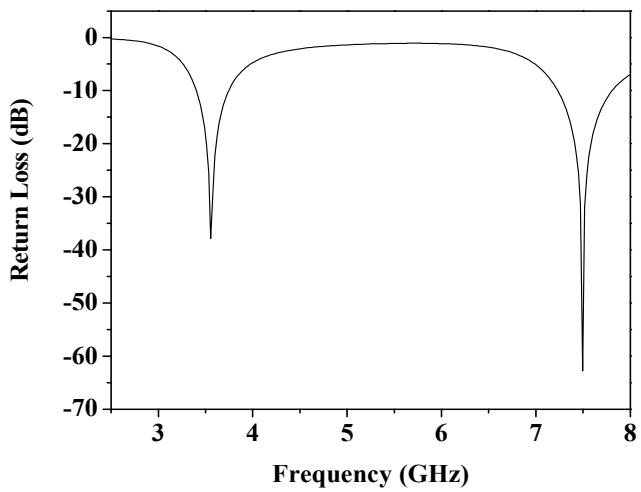


Figure 2. Return loss variation of the proposed antennas

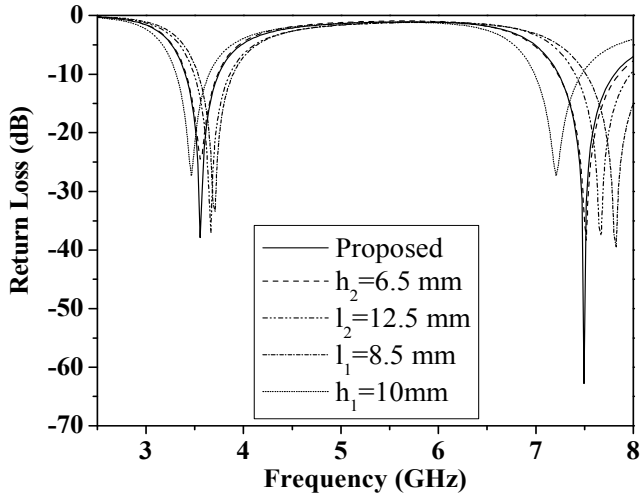


Figure 3. Return loss variation with length of proposed antenna

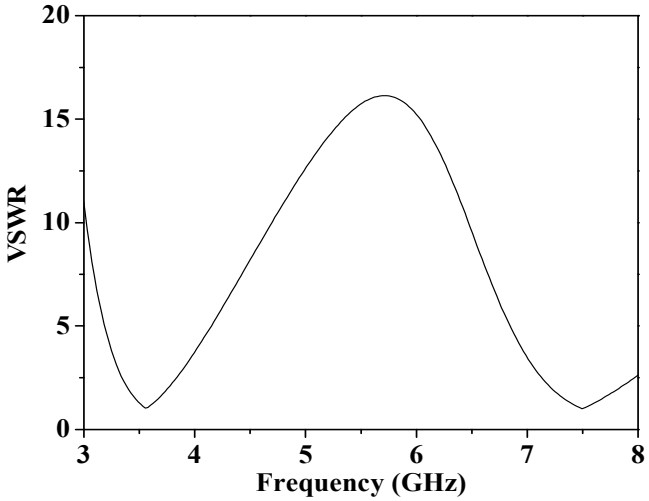


Figure 4. Variation of VSWR with frequency of proposed antenna

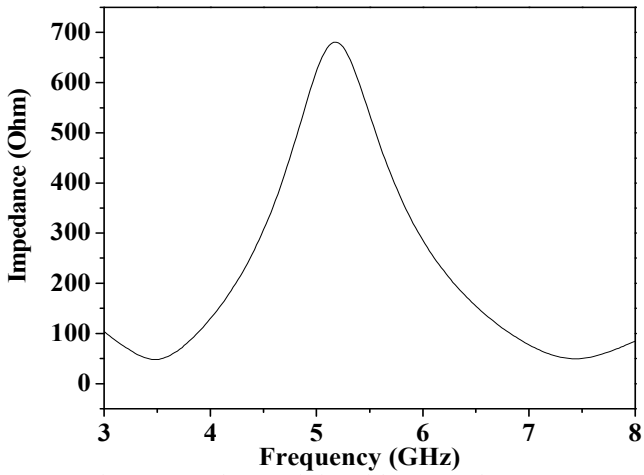


Figure 5. Impedance variation of the proposed antenna

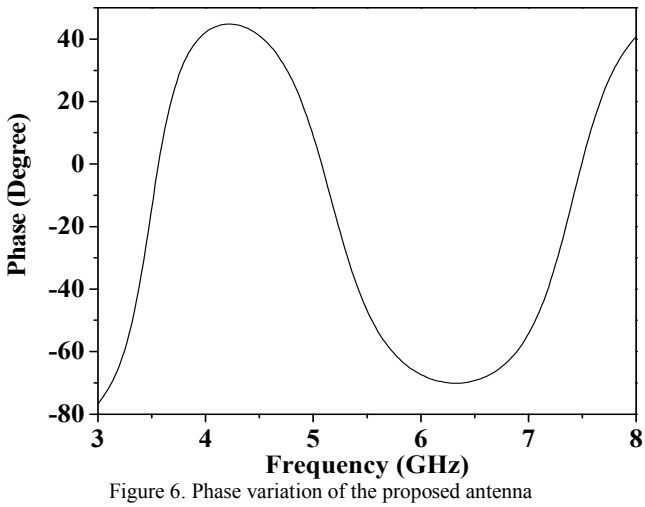


Figure 6. Phase variation of the proposed antenna

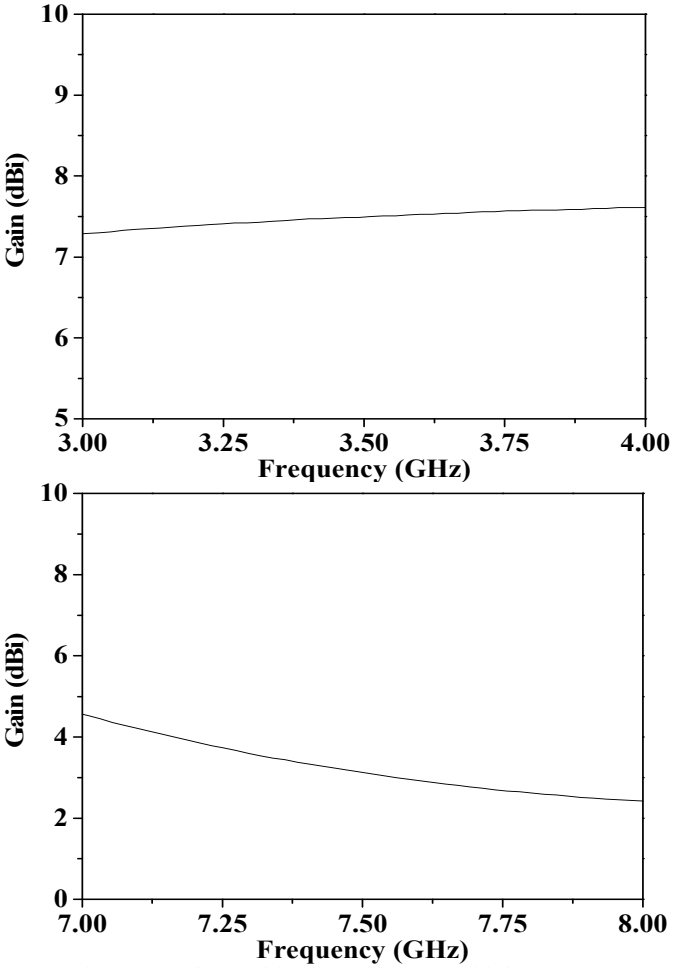


Figure 7. Total gain of the proposed antenna with frequency

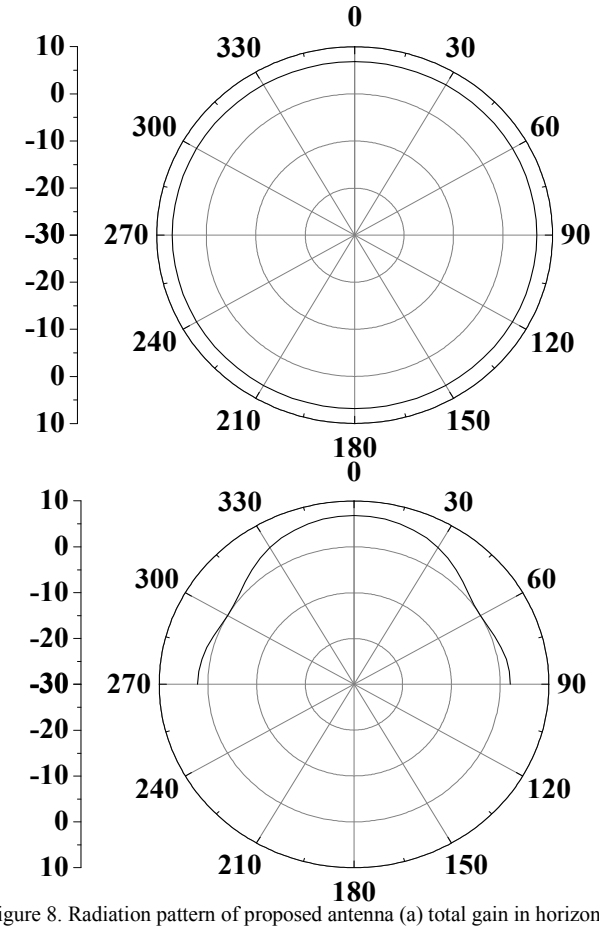


Figure 8. Radiation pattern of proposed antenna (a) total gain in horizontal plane (b) total gain in vertical plane and mirrored F-antenna (c) total gain in horizontal plane (d) total gain in vertical plane.

TABLE II. COMPARISON BETWEEN EXISTING ANTENNA AND OUR PROPOSED ANTENNA

Antenna	**Gain (dBi)**
	3.5 GHz WiMAX
Loaded U-antenna (Proposed)	7.51
Printed Multiband Antenna [3]	0.78
Capacitively Fed Hybrid Chip Antenna [4]	2.6-4.6
Wideband Printed Monopole Antenna [5]	2.1
Composite Monopole Antenna [6]	3.2-3.7
Triangle-Shaped Monopole antenna [7]	2.54

IV. CONCLUSION

A very compact structured dual frequency loaded U-antenna have been proposed. The antenna has a dimension of 23.5×14 (mm^2) and obtained operating bandwidths can cover the 3.5 GHz WiMAX band and the 7.25 GHz military satellite communication band. The proposed antenna can be embedded within the different portable devices employing WiMAX applications as it occupies small area. Sufficient bandwidth, high gain, and omnidirectional radiation characteristics of the proposed antenna reflects its effectiveness in mobile and wireless communication applications.

REFERENCES

[1] G. P. Karakoussis, A. I. Kostaridis, C. G. Biniaris and D. I. Kaklamani, "A Dual-Band Inverted-F Antenna Printed on a PC Card for the ISM and UNNI Bands", 0-7803 7700 1/03/$17.00 (C) 2003 IEEE.

[2] L. Pazin, N. Telzhensky, and Y. Leviatan, "Multiband Flat-Plate Inverted-F Antenna for Wi- Fi/WiMAX Operation," IEEE antennas and wireless propagation letters, Vol. 7, 2008.

[3] Shih-Ying Sun, Sheng-Yi Huang, and Jwo-Shiun Sun, "A Printed Multiband Antenna for Cellphone Applications", Microwave and Optical Technology Letters / Vol. 51, No. 3, March 2009.

[4] Peng-Yu Lai and Kin-Lu Wong, "Capacitively Fed Hybrid Monopole/Slot Chip Antenna for 2.5/3.5/5.5 GHz WiMAX Operation in the Mobile Phone", Microwave and Optical Technology Letters / Vol. 50, No. 10, October 2008.

[5] Chien-Yuan Pan, Tzyy-Sheng Horng, Wen-Shan Chen, and Chien-Hsiang Huang, "Dual Wideband Printed Monopole Antenna for WLAN/WiMAX Applications", IEEE Antennas and Wireless Propagation Letters, Vol. 6, 2007

[6] Kin-Lu Wong and Liang-Che Chou, "Internal Composite Monopole Antenna for WLAN/WiMAX Operation in a Laptop Computer", Microwave and Optical Technology Letters / Vol. 48, No. 5, May 2006.

[7] Y. Song, Y.-C. Jiao, G. Zhao, and F.-S. Zhang, "Multiband CPW-Fed Triangle-Shaped Monopole antenna for Wireless Applications", Progress In Electromagnetics Research, PIER 70, 329–336, 2007.

Automated Detection of Optic Disc and Blood Vessel in Retinal Image Using Morphological, Edge Detection and Feature Extraction Technique

Niluthpol Chowdhury Mithun, Sourav Das
Samsung R&D Institute Bangladesh
Dhaka-1205, Bangladesh
Email: {n.chy, sourav.das}@samsung.com

Shaikh Anowarul Fattah
Department of Electrical and Electronic Engineering,
BUET, Dhaka-1000, Bangladesh
Email: fattah@eee.buet.ac.bd

***Abstract*—Reliable, fast and efficient optic disc localization and blood-vessel detection are the primary tasks in computer analyses of retinal image. Most of the existing algorithms suffer due to inconsistent image contrast, varying individual condition, noises and computational complexity. This paper presents an algorithm to automatically detect landmark features of retinal image, such as optic disc and blood vessel. First, optic disc and blood vessel pixels are detected from blue plane of the image. Then, using OD location the vessel pixels are connected. The detection scheme utilizes basic operations like edge detection, binary thresholding and morphological operation. This method was evaluated on standard retinal image databases, such as STARE and DRIVE. Experimental results demonstrate that the high accuracy achieved by the proposed method is comparable to that reported by the most accurate methods in literature, alongside a substantial reduction of execution time. Thus the method may provide a reliable solution in automatic mass screening and diagnosis of the retinal diseases.**

Keywords—diabetic retinopathy; fundus image; retina; optic disc; blood vessel; morphological operation; feature extraction

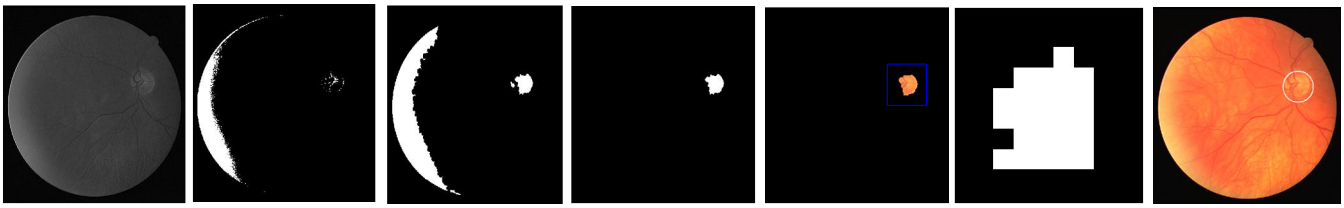
I. INTRODUCTION

Retinal imaging has become a potential area of research during the last 60 years. It is very useful in clinical care and management of patients with retinal as well as systemic diseases. One of this systemic diseases is Diabetic Retinopathy which occurs when the retina is been affected as a result of diabetes. Diabetes can cause weakening in the body's susceptible blood vessels and which results in a series of negative changes. This condition can lead to partial or even complete vision loss. Early detection and treatment of these diseases are crucial to avoid this preventable vision loss. Computer based retinal image analysis has been worked out since 1974 [1] and becoming very popular for quick detection of retinal diseases.

The identification of retinal features such as the optic disc and the retinal vessels is a prerequisite before systems can achieve more complex tasks identifying pathological entities. The location of the optic disc is crucial in retinal image analysis for localization of anatomical components in retina and measuring healthiness in retinal image. Moreover, it helps to track blood vessels. Optic disc is a cluster of high intensity pixels, with high grey-scale value [2]. A number of techniques exist for identification of optic disc in retinal photographs. In [3], Principal Component Analysis (PCA) has been used for localizing the original optic disc from noisy backgrounds. In [4], OD is localized by means of template matching and a deformable contour model was selected for its segmentation. Fuzzy convergence algorithm was investigated to detect optic disc in [5]. A new template-based methodology is proposed in [6], which uses morphological and edge detection techniques followed by circular hough transform. A fast technique to localize the OD is presented in [7], which is based on obtaining two projections of certain image features that encode the x and y coordinates of the OD. The location of the OD is then determined through searching the 1-D projections.

Blood vessel is another landmark features in retinal image for detecting retinal vein occlusion and grading the tortuosity for hypertension. Existing retinal vessel segmentation approaches can be classified into three categories: unsupervised methods, supervised learning methods and pathological retina based methods. In [8], a template matching based unsupervised approach is proposed to measure the degree of similarity between a given elongated inverse Gaussian template and the vessels. This method is enhanced by applying effective noise removal techniques in [9]. In [10], the supervised learning strategy is adopted for vessel segmentation, which yields very good results on healthy retinal images. However, many features, such as the Laplacian or Gaussian features require a high degree of smoothness in the non-vessel region. In [11], supervised method is adopted based on the Gabor features which work on the principle of detecting the low frequency elongated structure in retinal vessels intensity profile. In [12], 7-D vector composed of gray-level and moment invariants features has been computed for pixel representation and neural network has been used for pixel classification.

Most of the existing techniques suffer from uneven illumination, noises and computational load. In this paper, a novel methodology is presented for optic disc and blood-vessel detection in retinal image. The motivation was to develop a reliable method with significantly less computational complexity, whose performance is comparable to the best existing methods. The proposed method works in three major steps. First, optic disc is detected from blue plane of image using global thresholding, feature extraction and morphological operation. Secondly, blood-vessel pixel is detected using edge detection and NN based pixel classification. Finally, blood vessel is tracked by connecting vessel pixels by using optic disc as a reference length for measuring distances in retinal images.



(a) (b) (c) (d) (e) (f) (g)

Fig. 1. Illustration of the process for the calculation of the OD boundary approximation: a) blue plane image, b) – d) successive steps with thresholding, morphological operation and feature extraction for OD localization e) – f) postprocessing stages with candidate region selection on subimage from step [d], j) finally localized optic disc on original image

II. Proposed Optic Disc Detector

The optic disc (OD) is usually the brightest portion of retinal image. From a given RGB retinal image, unlike conventional approaches, in this paper, the input image is obtained by extracting only the blue plane. The reason behind selecting the blue plane image among the three planes (red, green and blue) is its comparatively weaker dynamic range, which makes the brightest portion (i.e. OD) significantly visible. It is to be mentioned that our experiments also support the idea that the blue plane image provides best result for optic disc detection. In the proposed method, two major steps are implemented for detecting the OD region from a given retinal image. First, optic disc is localized using global thresholding and feature extraction. Next, form the location obtained in first step, edge detection and morphological operation are performed to obtain the boundary region of the OD. The detailed steps are shown in Fig. 1.

A. Preprocessing by Global Thresholding

At first, for the blue plane image a global threshold is determined by using most commonly used Otsu method [16]. This threshold is scaled to a higher value and used to transform the image into a binary one. For a disease free image with perfect illumination condition, most of the bright pixels stay inside OD. Hence a crude estimate of bright OD region is obtained with so many disjoint bright pixels. In order to obtain smooth well connected OD region, some morphological operations are proposed. In this regard, first the erosion morphological operation is performed considering a square object Sμ, where μ×μ(μ>1). Then dilation operation followed by again erosion operation is performed with a disc object Dλ, where λ(λ>1). This will connect bright pixels and some blobs will be formed. The choice of λ and μ may be dependent on the illumination condition, resolution, and size of image. From detailed experimentation it is found that the values of μ and λ can be chosen as 3 and 5, respectively.

B. Feature Extraction and OD Detection

The features need to be chosen in such a way that they are sufficiently distinctive and less sensitive to environment. With a view to localize the OD, following 4 features are extracted.

- Area(A): This is the total number of pixels of the Blob.
- Ratio of major axis and minor axis (R): The length of the major axis and minor axis are obtained from the ellipse that fits the Blob best.
- Compactness(C): This feature determines the closeness to circular shape of the segmented region of retinal image, which is evaluated as

$$Compactness = \frac{Area}{Sqr(Perimeter)}$$

Then, small areas created by morphological operation but don't have a sufficient density of bright pixels are removed. As the OD is of circular shape, the ratio of major axis to minor axis (R) of blob is expected to be a small value and compactness (C) to be close to unity. For all the remaining blobs, Rand C is calculated. The blob with the minimum R and maximum C is selected as the optic disc region. However, if the major axis value is more than three times of minor axis value for the detected blob, the blob is rejected as optic disc. For such a case, the total process is started again by selecting a smaller threshold than the previous one and this threshold is used to transform the image into a binary one. Now again morphological operation is performed as previous and the features are extracted. Again the ratio R is calculated and checked if it is less than 3. If the criterion is satisfied, the blob is considered to be representing optic disc. In brief the three conditions are:

$\Upsilon 1: A > THA$[Threshold is selected based on image size]

$\Upsilon 2: R < THR$[Threshold is selected as 3]

$\Upsilon 3: THC < C < 1$ [Threshold is selected as 0.8]

Hence, the Blob selection depends on the chosen threshold values THA, THR, and THC for area, ratio, and compactness, respectively and finally it is selected when condition Υ is true.

$$\Upsilon: \Upsilon 1 \cap \Upsilon 2 \cap \Upsilon 3$$

Next, other small objects present in the image are removed. Finally, morphological closing operation is performed with a disc of diameter 20 on the remaining blob. This will complete optic disc localization procedure.

C. Post-processing

In order to obtain better segment of OD region, a sub-image is taken from the previously segmented OD region from the blue plane of image, whose area is extended to 80×80 outside the detected region. Now median filtering is employed in the sub-image to remove small noises. Next 10×10 smaller non-overlapping blocks of the sub-image are analyzed. From the summation of pixel values of all the small blocks, mean pixel area is calculated. The regions having area more than the mean are transformed completely into white and the other regions are transformed into black. Finally after filling the holes and removing small areas, a better segmented OD region is obtained. From the estimated bounding box of the OD, a circle's center and radius is calculated. This circle indicates boundary of the OD region.

III. Proposed Blood vessel Segmentation Method

For blood vessel detection, edge detection algorithm is widely used. It can be classified into two categories, namely edge enhancement [17] and edge fitting [18]. The edge enhancement method is susceptible to noise whereas the edge fitting method suffers from the burden of repeated and large computations. In this paper, a new edge-finding method is proposed, which is designed especially for blood vessel

detection. The following stages may be identified for vessel segmentation: 1) preprocessing for gray-level homogenization and vessel enhancement, 2) feature extraction through edge finding and 3) post-processing for filling pixel gaps in detected blood vessels and removing falsely-detected pixels. Note that, unlike the OD detection method, in blood vessel detection, green plane of the original RGB retinal image is considered, which offers the highest contrast and in the green plane, blood vessels are best represented.

A. Preprocessing

Fundus images often show significant lighting variations, poor contrast and noise. In order to reduce these imperfections, a preprocessing stage is needed. At first, the workable region is extracted via mask provided by the corresponding database. Then, background homogenization is necessary as background pixels may have different intensity for the same image and, although their gray-levels are usually higher than those of vessel pixels, the intensity values of some background pixels are comparable to that of brighter vessel pixels. Background homogenization has been performed through filtering the high frequency component with the help of moving average method.

B. Feature Extraction through Edge Detection

The aim of the feature extraction stage is representation of pixel in terms of some quantifiable measurements which may be used in the classification stage to decide whether the pixel is of a real blood vessel. In this paper, a new edge finding process was applied for this purpose. The blood vessel are darker compared to the surrounding environment. Moreover the blood vessel diameter is fixed for a selected database (5~10 pixels for STARE and DRIVE database). The detected size of OD in Section-II can be used to automatically detect vessel diameter. Since the spatial width of the blood vessel is no more than 5~10 pixels, whenever an edge of any blood vessel is found, it can be assumed that moving in the same direction through the blood vessel region, the other edge of the blood vessel can be found within next 5~10 pixels. The problem arises, when searching in a specific direction fails to detect the edge. To solve the problem and identify all blood vessels, edge-searching is done in possible four directions, such as: horizontally from left to right, vertically from top to bottom, diagonally towards the top right corner and diagonally towards the left-bottom corner. Finally, result from the four directions is merged to get the approximate path of blood vessel. Fig 2 describes the procedure of edge-finding.

C. Post-processing

The post-processing stage is aimed at filling pixel gaps in detected blood vessels and removing falsely detected isolated vessel pixels. From visual inspection of the NN output, vessels may have a few gaps (i.e., pixels completely surrounded by vessel points, but not labeled as vessel pixels). To overcome this problem, a iterative filling operation is performed by considering that pixels with at least six neighbors are classified as vessel pixel; is also a vessel pixels. Besides, small isolated regions miss-classified as blood vessel pixels is also observed. Difficulties arise in connecting the correlated pixels, as different blood vessels may overlap in 2-D image representation. Some artifacts remain in our detected image because of this. The post processing part will be further researched in future study to develop a more efficient connection algorithm and better result.

IV. Results and Analysis

A. Materials

To evaluate the methodology, two publicly available databases, such as the DRIVE [14] and STARE [15] are used. Apart from being public; they provide manual segmentations for performance evaluation. The DRIVE database comprises 40 eye-fundus color images. The database is divided into two sets: a test set and a training set, each of them containing 20 images. The test and training set provides the corresponding FOV masks for the images and manual segmentations generated by specialists for each image. The STARE database, originally collected by Hoover, comprises 20 eye-fundus color images. The database contains two sets of manual segmentations made by two different observers.

B. OD Detection and Blood Vessel Segmentation

Performance is evaluated by measuring the degree of overlap between the true OD regions in 'manually leveled' images and the approximated regions obtained with the described approach. The proposal by Lalonde et al.[18] was used with this purpose: an overlapping score S is defined to measure the common area between a true OD region T and a detected region D as

$$S = \frac{Area\,(T \cap D)}{Area(T \cup D)}$$

The proposed algorithm was applied to calculate overlapping constant S for all the images of DRIVE and STARE database.

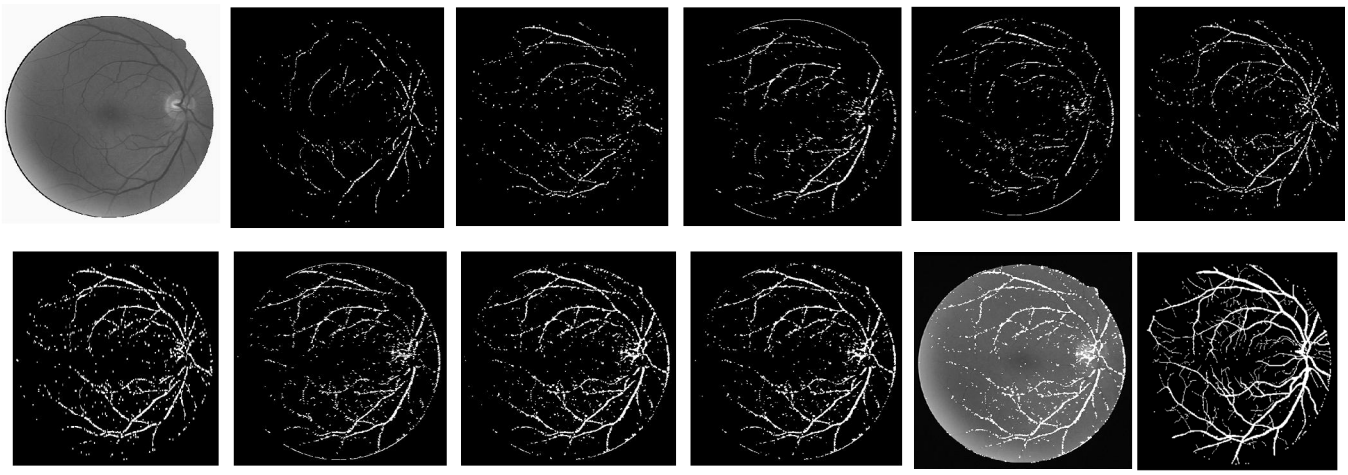


Fig.2. Illustration of process for detection of blood vessel. a) The green component of the original image, b)-e) successive steps of edge finding in horizontal direction, in vertical direction, diagonally towards from bottom-left to top-right and diagonally towards bottom-right f) merged result of (b) and (c), g) merged result of (d) and (e) h) merged result of edge-finding i) overall merged result after connecting the adjacent components , j) NN network output and filtered resultant image , k) superimposed on the original image l) finally identified vessel pixels after post-processing

TABLE 1: Performance of The Proposed OD Method in terms of Success Rate and Computational Time

Database	STARE	DRIVE	TOTAL
Normal Images	31	33	64
Diseased Images	50	7	57
Number of Images	81	40	121
Success Rate (%)	91.3	100	94.2
Average Computation Time (seconds)	1.45	1.28	

In our experiment, S value greater than.0.75 is selected as successfully detected. The accuracy of the proposed OD detection method are summarized in Table 1. It is observed from the table that, the proposed OD detection method provides a high level of accuracy. For well captured image, like DRIVE databases, this method achieves 100% accuracy. This proves reliability of using this technique in fundus image captured with well-focused fundus camera.

However, in order to quantify the algorithmic performance of the proposed vessel detection method on a fundus image, the resulting segmentation is compared to its corresponding gold-standard image. This image is obtained by manual creation of a vessel mask in which all vessel pixels are set to one and all non-vessel pixels are set to zero. Thus, automated vessel segmentation performance can be assessed.

TABLE 2 : CONTINGENCY VESSEL CLASSIFICATION TABLE

	Vessel Present	Vessel Absent
Vessel detected	True Positive (TP)	True Negative (NP)
Vessel not detected	False Positive (FP)	False Negative (FP)

The proposed vessel segmentation algorithm was evaluated in terms of sensitivity (Se), specificity (Sp), positive predictive value (Ppv), negative predictive value (Npv), and accuracy (Acc) as in [12]. Taking above table into account, these metrics are defined as-

$$Se = TP / (TP + TN)$$
$$Sp = TN / (TN + FN)$$
$$Ppv = TP / (TP + FP)$$
$$Npv = TN / (TN + FN)$$
$$Acc = (TP + TN) / (TP + FP + TN + FN)$$

Se and Sp metrics are the ratio of well-classified vessel and nonvessel pixels, respectively. Ppv is the ratio of pixels classified as vessel pixels that are correctly classified. Npv is the ratio of pixels classified as background pixel that are correctly classified. Finally, Acc is a global measure providing the ratio of total well-classified pixels. Se, Sp, Ppv, Npv and Acc values are computed for each image considering FOV pixels only. The results are listed in Table 3.

The last row of the table 3 shows average values for all the images in this database. Here for edge detection, threshold values need to be chosen considering the facts that lowering the threshold value generates continuous blood vessel with the tradeoff of noisy components. On the other hand, increasing the threshold values generates discrete blood vessel components but less noise. Examining the whole database these threshold values were chosen to maximize the output as well as reducing the noise.

TABLE 3 : VESSEL SEGMENTATION PERFORMANCE RESULTS

Image	Se	Sp	Ppv	Npv	Acc
1	0.5682	0.9668	0.626	0.9581	0.9312
2	0.5878	0.9602	0.6274	0.9533	0.942
3	0.3862	0.986	0.6935	0.9258	0.9362
4	0.4789	0.9734	0.6457	0.9486	0.9279
5	0.401	0.9791	0.6653	0.9405	0.9425
6	0.4102	0.9791	0.6794	0.939	0.9238
7	0.4405	0.9672	0.5749	0.945	0.9391
8	0.3594	0.9772	0.5976	0.9419	0.9241
9	0.3316	0.9845	0.6538	0.9435	0.9316
10	0.4457	0.9754	0.6194	0.9515	0.9418
11	0.5558	0.9532	0.5386	0.9562	0.9476
12	0.5102	0.9652	0.5806	0.9542	0.9259
13	0.4593	0.9684	0.6118	0.9429	0.9386
14	0.5183	0.966	0.5725	0.958	0.9398
15	0.5638	0.9525	0.4778	0.9659	0.9247
16	0.5008	0.9663	0.5958	0.9512	0.9443
17	0.4699	0.9743	0.6277	0.9522	0.9317
18	0.5233	0.9667	0.575	0.9593	0.9416
19	0.5774	0.9646	0.5963	0.9619	0.9325
20	0.422	0.9766	0.5888	0.9551	0.9358
Average	0.4705	0.9701	0.6074	0.9502	0.9363

In order to compare our approach to other retinal vessel segmentation algorithms, average Acc is used. Table 4 shows performance comparison in terms of average Acc with several published methods: Chaudhuri et al. [8], Staal et al. [10], Soares et al. [11], and Cinsdikici and Aydin [13]. From the table, it is found that the results of proposed method are comparable to that of other detection techniques.

TABLE 4: PERFORMANCE RESULTS COMPARISON IN TERMS OF AVERAGE ACCURACY

Method Type	Method	Result (Acc)
Supervised	Staal et al. [10]	0.9441
	Soares et al. [11]	0.9466
	Proposed Method	**0.9363**
Unsupervised methods	Chaudhuri et al. [8]	0.8773
	Cinsdikici and Aydm [13]	0.9293

V. CONCLUSION

The everyday cost associated with eye care provider's decisions and the ever-increasing numbers of retinal images are the major motivations for the adoption of computer based retinal image analysis. In this paper, a reliable yet simple methodology for optic disc and blood vessel detection from retinal image is presented. The results show that the proposed method is robust to variation of illumination and image size. For well captured image with good resolution, it provides significant level of accuracy. Thus, the proposed method may be highly effective in designing a reliable computer based retinal image analysis system, as now a days the imaging quality is high and the amount of data is also high.

REFERENCES

[1] Yamamoto S. and Yokouchi H.,"Automatic diagnosis of color fundus photographs," in Proc. Int. Conf. Pattern Recognition, 1974, pp.268–269.

[2] Patton N., Aslam T.M., MacGillivray T., Deary I.J. , Dhillon B., Eikelboom R.H. , Yogesan K. , Constable I.J. , "Retinal image analysis: Concepts, applications and potential", Progress in Retinal and Eye Research 25, 2006, pp.99–127.

[3] Li,H.,Chutatape,O., "Automated feature extraction in color retinal images by a model based approach.", IEEE Trans. Biomed. Eng., vol.51,2004,pp.246–254.

[4] Lowell J., Hunter A., Steel D.,Basu A.,Ryder R., Fletcher E., Kennedy L., "Optic nerve head segmentation," IEEE Trans. Med. Imag.vol. 23, no. 2,2004 pp. 256–264.

[5] Hoover,A.,Goldbaum,M., "Locating the optic nerve in a retinal image using the fuzzy convergence of the blood vessels.", IEEE Trans. Biomed. Eng. Vol.22,2004, pp.951–958.

[6] Aquino A., Gegúndez-Arias M.E., and Marín D.,"Detecting the Optic Disc Boundary in Digital Fundus Images Using Morphological, Edge Detection, and Feature Extraction Techniques", IEEE Trans. on Med. Imaging, vol. 29, no. 11, 2010, pp. 1860-1869.

[7] Mahfouz, A.E., Fahmy, A.S., "Fast Localization of the Optic Disc Using Projection of Image Features", IEEE Trans. Imag. Proc., vol. 19, no. 12, 2010, pp. 3285-3289.

[8] Chaudhuri, S.,Chatterjee,S., Katz,N. Nelson,M. and Goldbaum, M. "Detection of blood vessels in retinal images using two-dimensional matched filters," IEEE Trans. Med. Imag., vol. 8, no. 3, Sep. 1989, , pp. 263–269.

[9] Hoover, A.. Kouznetsova,V. and Goldbaum, M., "Locating blood vessels in retinal images by piecewise threshold probing of attached filter response," IEEE Trans. Med. Imag., vol. 19, no. 3, Mar. 2000, pp. 203–210.

[10] Staal, J., Abràmoff, M. D., Niemeijer,M. Viergever,M. A. and Ginneken,B. v. "Ridge based vessel segmentation in color images of the retina," IEEE Trans.Med. Imag., vol. 23, no. 4, ,Apr. 2004, pp. 501–509.

[11] Soares, J. V. B. , Leandro, J. J. G., Cesar, R. M., Jelinek, Jr, H. F. and Cree, M.J. "Retinal vessel segmentation using the 2-D Gabor wavelet and supervised classification," IEEE Trans. Med. Imag., vol. 25, no. 9, Sep. 2006, pp.1214–1222.

[12] Marín D., Aquino A., Gegúndez-Arias M. E, and Bravo J. M., "A New Supervised Method for Blood Vessel Segmentation in Retinal Images by Using Gray-Level and Moment Invariants-Based Features", IEEE Trans. Med. Imag., vol. 30, no. 1, Jan. 2011, pp. 146-157.

[13] Cinsdikici, M. G. and Aydin D., "Detection of blood vessels in ophthalmoscope images using MF/ant (matched filter/ant colony) algorithm," Comput. Methods Programs Biomed., vol. 96, 2009, pp. 85–95.

[14] Research Section, Digital Retinal Image for Vessel Extraction (DRIVE) Database. Utrecht, The Netherlands, Univ. Med. Center Utrecht, Image Sci. Inst. [Online]. Available: http://www.isi.uu.nl/Research/Databases/DRIVE.

[15] STARE ProjectWebsite. Clemson, SC, Clemson Univ. [Online]. Available: http://www.ces.clemson.edu/

[16] Otsu N., "A threshold selection method from gray-level histograms". IEEE Trans. Sys., Man., Cyber.9(1),1979,:pp.62–66

[17] Torre V. and Poggio T.A. "On Edge Detection" IEEE Trans. Patt. Anal. Machine Intell. vol. PAMI-8, Mar. 1986; pp. 147-163,

[18] Lalonde M., Beaulieu M., and Gagnon L., "Fast and robust optic disc detection using pyramidal decomposition and Hausdorff-based template matching," IEEE Trans. Med. Imag., vol. 20, no. 11, Nov. 2001; pp.1193–1200,.

[19] Mithun N. C., Rashid N. U. and Rahman S. M. M., "Detection and Classification of Vehicles From Video Using Multiple Time-Spatial Images,"in IEEE Trans. Intell. Transp. Syst., vol. 13, no. 3, Sep. 2012,pp. 1215–1225.

Selection of Proper Frequency Band and Compatible Features for Left and Right Hand Movement from EEG Signal Analysis

Manjurul Ahsan Riheen, Md. Wasiur Rahman, A. B. M. Aowlad Hossain
Department of Electronics and Communication Engineering
Khulna University of Engineering & Technology
Khulna, Bangladesh
r_ahsan@hotmail.com, wasi09ece@live.com, aowlad0403@yahoo.com

***Abstract*— Electroencephalogram (EEG) signal plays an important role in the field of brain-computer interface (BCI) which has diverse applications ranging from medicine to entertainment. BCI acquires brain signals, extracts informative features and translates these features into a control signal for an external device. The Purpose of this work is to select proper frequency band and to extract suitable features for left and right hand movements from EEG signals analysis. The Discrete Wavelet Transform (DWT) is used to extract different significant features, which separates Alpha, Beta and Theta band of frequencies of the EEG signal. Extracted EEG features of different bands are classified using an artificial neural network (ANN) trained with the back propagation (BP) algorithm. The classification rate shows that Beta band (97.5%) has higher mapping precision and better convergence rate than the other bands, alpha (93.2%) and theta (87.8%). Finally an ANN self-organizing feature mapping (SOFM) is used to find the compatible feature for EEG bands related to hand movement. SOFM analysis shows that approximate entropy (ApEn) for theta band and scale variance for alpha and beta band can be used as compatible feature. The results of this study are expected to be helpful in brain computer interfacing.**

Keywords*— *Electroencephalogram, Discrete Wavelet Transform, Feature Extraction, Artificial Neural Network, Self-Organizing Feature Mapping

I. INTRODUCTION

In recent years, stroke, traumatic brain injury (TBI) and spinal cord injury (SCI) are the leading causes of disability affecting over a million individuals annually [1]. A large number of people have reported severe trouble with hand function [2]. Failure of hand function causes severe problems in leading of life for affected persons [3]. In certain situation, BCIs is the new hope of treatment which is in research state. EEG signal is the most popular way of interpreting the brain activities in the realm of non-invasive BCI. They are well-suited and there is evident that they can be used in artificial hand movements [4].

EEG gives the graphical representation of brain electrical activities. The electrical activities of the brain are recorded from electrodes located on the scalp. EEG have certain bands which are distinguished by their different frequency ranges [5]: Theta waves varies in the range of 4 Hz to 7 Hz and its amplitude generally more than 20 μV, Alpha wave varies within the range of 8 to 13 Hz and 30-50 μV amplitude. For Beta wave, the frequencies vary between 13 Hz to 30 Hz, and usually have a low voltage between 5-30 μV. Beta wave part of EEG signal usually related to dynamic thinking, concentration etc. EEG signals are extensively studied by numerous researches to classify different mental or brain activities [6]-[9]. Selection of appropriate and compatible feature is very crucial for effective classification. In this study, we plan to find proper frequency band and compatible feature in each band for left and right hand movement using wavelet transform and neural network based EEG signal analysis.

Wavelet provides a consolidated system for different techniques that used in the biomedical signal processing which developed for various applications [10]. It is a robust technique to represent the signal in time-frequency domain. The capability of time-frequency domain analysis of wavelet transform can be useful to separate the alpha, beta and theta band and consequently to extract important features.

Numerous techniques associated with signal processing have been proposed to extract the desired features classification. Artificial Neural networks have been applied in several studies for EEG analysis. ANNs is executed through training algorithms with specified learning criteria to imitate the learning mechanisms of biological neurons [11], [12]. Different types and structures of neural networks are available. In this paper, we have considered the feed forward back propagation neural network for classify the alpha, beta and theta band features. Furthermore, self-organizing feature maps can be used classify based on the information on how they are grouped in the input space. They fluctuate from competitive layers in those neighboring neurons. Self-organizing map learn to distinguish adjacent parts of the input space. Thus, self-organizing maps are gaining knowledge from both the distribution and topology of the input vectors.

The paper is organized as follows: the next section describes the proposed methodology which deals with wavelet decomposition of EEG bands, feature extraction and ANN

based classification. Section III describes the data acquisition and result analysis followed by conclusion in Section IV.

II. Proposed Methodology

The purpose of this study has two fold; first to select appropriate band and then to find compatible feature from various features extracted from EEG signals. We used discrete wavelet transform to select different band in order to compute different features for the bands. Finally, an ANN trained with back-propagation algorithm used to select the best band and a SOFM is used to find the compatible features taking the different extracted feature as input. The block of proposed methodology is shown in Fig. 1:

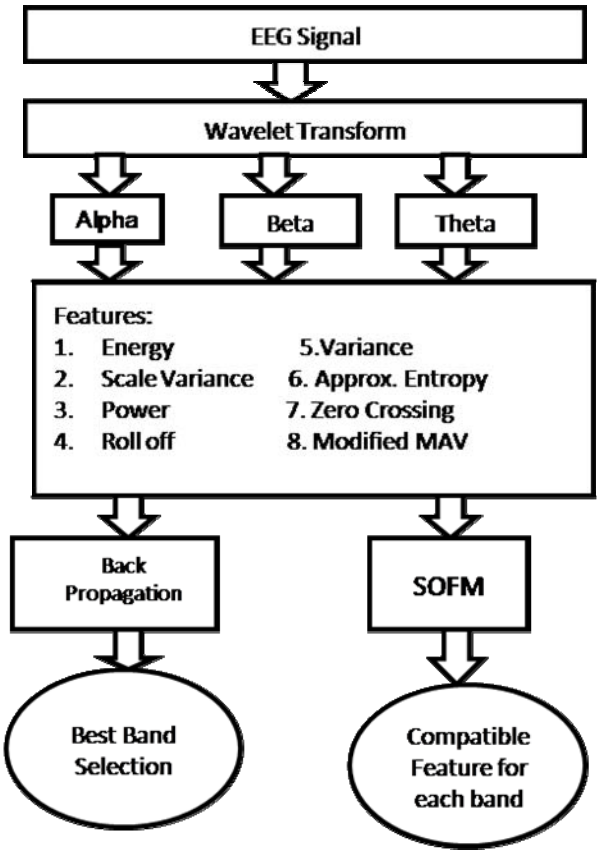


Fig. 1. Block diagram of proposed methodology

A. Feature Extraction

Transforming the input data into the set of features is feature extraction. A few number of feature extraction technique is proposed in the literature to represent the EEG signal, such as power spectra, adaptive autoregressive (AAR) and wavelet transform. A great advantage of wavelet transform based analysis is that it can represent the signal in both time and frequency domain which is better compared to the Fourier transform and short time Fourier transform.

Wavelet analysis is splitting up of a signal in shifted and scaled form of the original wavelet. Wavelet is a small waveform of effectively brief duration that has an average value of zero. It conserves the time-frequency components of a signal at different resolution and scales. Wavelets can be used as filter with multi resolution advantage. The resolution of the signal is determined by the filtering operations and the scale is determined by up sampling and down sampling operations.

As a result, the discrete wavelet transform DWT can be achieved by successive low pass and high pass filter at discrete time domain is shown in Fig. 2. This is called the Mallat algorithm which was proposed by Mallat [13] where x[n] is the input signal which passes through a high pass filter where the impulse response is h[n] and simultaneously passes through the low pass filter with an impulse response g[n]. The output of the high pass filter provides the detail coefficient and the output of the low pass filter provides the approximate coefficient. The filter output is given in (1) and (2) where k varies from -∞ to ∞ [14].

$$Y_{high} = \sum x[k]\ h[2n-k] \tag{1}$$

$$Y_{low} = \sum x[k]\ g[2n-k] \tag{2}$$

We can decompose the EEG signal in both time and frequency domain using DWT and then different features can be computed. There are numbers of time and frequency domain features used in several studies, among them we have considered eight different features which are given below. Here, N indicates the no. of samples. For wavelet decomposition, we used db8 wavelet up to level four to six. Fourth, fifth and sixth level coefficients of wavelet decomposition corresponds the beta, alpha and theta band of frequencies respectively.

1. <u>*Energy:*</u>

Energy of a signal can be defined as a simple square integral. It can be used as a feature of EEG signal. It can be expressed as-

$$Energy = \sum_{n=1}^{N} \left| x_n^2 \right| \tag{3}$$

2. <u>*Scale Variance (SV):*</u>

Scale variance known as a log- variance of wavelet coefficient. It can be expressed as-

$$SV = \log\left(\mathrm{var}(x_n)\right) / \log 2 \tag{4}$$

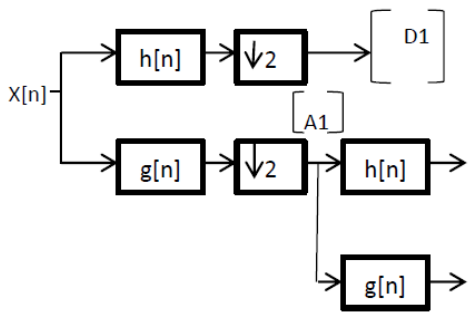


Fig. 2. Wavelet Decomposition Tree

3. Power:

Power can also be used as a feature. It can be expressed as-

$$Power = \sqrt{\frac{\sum_{n=1}^{N} x_n}{N}} \quad (5)$$

4. Rolloff:

The rolloff can be defined as the frequency below which 85% of the magnitude distribution of the spectrum is intensed. It is also a measure of spectral shape which can be written as-

$$R = 0.85 \times \sum_{n=1}^{n/2} |x_n| \quad (6)$$

5. Variance:

Variance is the mean value of the square of the deviation of the signal. However, mean of EEG signal is close to zero. Hence, we calculate the variance of EEG following the equation below-

$$VAR = \frac{1}{N-1} \sum_{n=1}^{N} x_n^{\,2} \quad (7)$$

6. Approximate Entropy (ApEn):

Approximate entropy can be defined as a measure of randomness or regularity [15]. It represents the information in a time series. A low value of the entropy signifies that the time series is deterministic; on the contrary, a high value signifies the randomness. ApEn has been used to characterize different biomedical signals, such as EEG, ECG.

If N data points form a time series $\{X(n)\} = X(1), X(2), X(3),\ldots, X(N)$, one should follow these steps to compute ApEn [15]:

i. Form N-m+1 vectors $X(1), X(2),\ldots,X(N-m+1)$ defined by:

$$X(i) = [\, X(i), X(i+1), \ldots, X(i+m-1)],\ i=1,2,\ldots,n-m+1$$

ii. Calculate the distance between $X(i)$ and $X(j)$,$d[X(i),X(j)]$, as the maximum absolute difference between their respective scalar components,

$$d[X(i),X(j)] = \max_{k=1,2,..m} \left| x(i+k-1) - x(j+k-1) \right| \quad (8)$$

iii. For a given $X(i)$, count the number of j, (j=1,2,…, N-m+1) so that $d[X(i),X(j)] \le r$, denoted as $N^m(i)$, i=1,2,....N-m+1

$$C_r^{\,m}(i) = N^m(i)/(N-m+1) \quad (9)$$

iv. Compute the natural logarithm of each $C_r^{\,m}(i)$ and average it over I,

$$\Phi^m(r) = \frac{1}{N-m+1} \times \sum_{i=1}^{N-m-1} \log\left(C_r^{\,m}(i)\right) \quad (10)$$

v. Increase the dimension m to m+1.Repeat steps i–iv to get $C_r^{\,m+1}(i)$ and $\Phi^{m+1}(r)$.

vi. ApEn can be calculated by the following formula:

$$\mathrm{ApEn}(m,r,N) = \Phi^m(r) - \Phi^{m+1}(r) \quad (11)$$

7. Zero Crossing:

Zero crossing (ZC) indicates the number of times that the amplitude value of EEG signal crosses the zero y-axis. It is an approximate estimation of frequency domain properties. It can be calculated as-

$$\mathrm{ZC} = \sum_{n=1}^{N-1} \mathrm{sgn}(x_n * x_{n-1}) \cap |x_n - x_{n-1}| \ge \text{Threshold} \quad (12)$$

$$\text{Where, } \mathrm{sgn}(x) = \begin{cases} 1, & x \ge \text{Threshold} \\ 0, & \text{Otherwise} \end{cases}$$

8. Modified Mean Absolute Value (MMAV):

Mean Absolute Value (MAV) is the moving average of full-wave rectified EEG. In other words, it is calculated by taking the average of absolute value of EEG signal. So, MMAV is the extension of MAV by using the weighting function w_n.

$$\mathrm{MMAV} = \frac{1}{N} \sum_{n=1}^{N} (w_n * |x_n|) \quad (13)$$

$$\text{Where, } w_n = \begin{cases} 1, & 0.25N \le n \le 0.75N \\ 0.5, & \text{Otherwise} \end{cases}$$

B. Classification

Artificial neural networks are computing systems those made up of large number of simple, very interrelated processing elements (nodes) that imitate the structure and operation of the biological nervous system [16]. Learning in ANNs is executed by training algorithms based on the learning mechanisms of biological neurons. They are various types of neural network varying fundamentally in the way they learn. In this paper, feed forward back propagation neural network is considered.

A feed forward back propagation neural network consists of a number of simple neurons like processing unit. Every unit in a layer is connected with all the units in the previous layer. But, these connections are not all equal each connection may have a different weight or strength. A BP network may contain two or more layers Output layer decides the solution of the problem. Another hidden processing layer is used to hold the nonlinearity and complexity of the problem. Selection of proper number of hidden layers is a vital task. The proposed neural network structure is shown in Fig. 3.

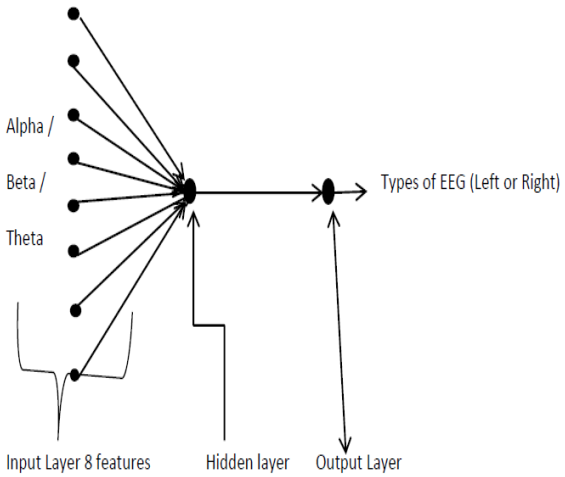


Fig. 3. Proposed Neural Network Training Structure

In this paper, we have considered 8 features for classification. So, neural network is designed with 8 input nodes as a train dataset and one output node. The network is trained with parameters a gradient descent algorithm. Performance of the network is specified as Mean Square Error (MSE). The training was stopped when the MSE between the network outputs and the targets was lesser than or equal to 0.00001. The learning rate is fixed at 0.05. The number of training epochs was fixed uniformly at 2000.

Furthermore, SOFM is one kind of unsupervised ANN. In this type of ANN the training of the network is totally drive by data and there are no target no target results [17]. It can be used for selecting features innate to the classification. In this paper, we have taken the benefit of this property of SOFM. The neurons in the layer of an SOFM are set in physical positions in accordance with a topology function. Hence, mapping is completed by the relative space between the points. Points that are near to each other in the input space are mapped to nearby map units in the SOFM. Here, we have considered 6*6 hexagonal topological function. The number of training epochs was fixed uniformly at 200.

III. Result Analysis

We have collected EEG data of left and right hand movement. The subject of the dataset is a 21 year old, right handed male with no known medical conditions. The EEG consists of actual random movements of left and right hand recorded with eyes closed. There are nineteen electrodes. The order of the electrodes is FP1 FP2 F3 F4 C3 C4 P3 P4 O1 O2 F7 F8 T3 T4 T5 T6 FZ CZ PZ. The length of the signal for each electrode is 3200 samples and sample rate is 500 Hz. The dataset was downloaded from the website [18]. Figure 4 shows the raw EEG data for left and right hand movement. The raw signals are then decomposed in different bands using discrete wavelet transform which are shown in Fig. 5. Selection of decomposition levels depends on Alpha, Beta and Theta frequency band. The DWT, back propagation trained ANN and SOFM are performed by using the MATLAB package.

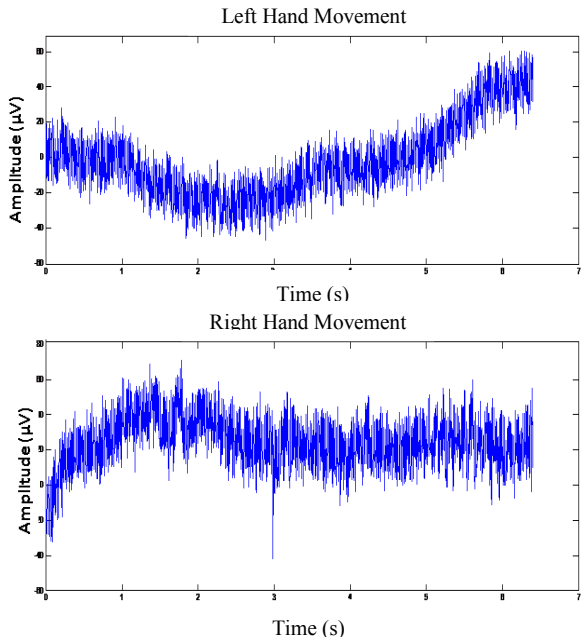


Fig. 4. Raw EEG data for left and right hand movement

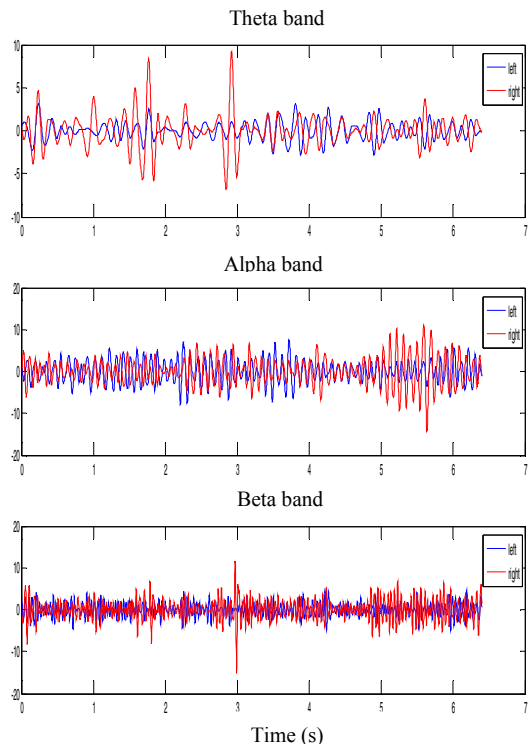


Fig. 5. Theta, Alpha, and Beta band separation from EEG

The performance of the classification of the EEG data is evaluated in terms of the three parameters i.e., Sensitivity (SE), Specificity (SP) and Accuracy.

Sensitivity (SE) indicates the capacity of correctly identified positive cases. It is also known as a true positive ratio.

$$SE = \frac{TP}{TP + FN}$$; TP=True Positive and FN = False Negative

Specifity (SP) indicates the capacity of correctly identified negative cases. It is also known as a true negative ratio.

$$SP = \frac{TN}{FP + TN}$$; TN= True Negative and FP= False Positive

Accuracy indicates that the proportion of correct classified instances. It can be calculated as-

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

Table 1 shows the number of correct and false classification of EEG dataset in different frequency bands. Table 2 shows the classification performance of different frequency band by considering sensitivity, specificity and overall accuracy. The sensitivity and specificity of theta, alpha and beta band are 83.8% and 92.9%, 100% and 88%, 95.2% and 100%, respectively. The overall accuracy of theta, alpha and beta band is 87.8%, 93.2% and 97.5%. From, the above table and performance parameter we can be concluded that beta band has higher convergence rate than the alpha and theta band.

TABLE I. CLASSIFICATION RESULTS OF DIFFERENT BAND

Hand	**Theta Band**		**Alpha Band**		**Beta Band**	
	No. of Classification		*No. of Classification*		*No. of Classification*	
	Correct	False	Correct	False	Correct	False
Left	15	4	19	0	18	1
Right	18	1	16	3	19	0

TABLE II. SENSITIVITY, SPECIFICITY AND ACCURACY OF THE CLASSIFICATION

	Sensitivity	**Specificity**	**Accuracy**
Theta	83.8%	92.9%	87.8%
Alpha	100%	88%	93.2%
Beta	95.2%	100%	97.5%

The features are then mapped with SOFM to find the compatible features of each band. Figure 6 shows the input feature mapping of frequency band theta, alpha and beta respectively by using SOFM. In this figure, lighter colors represent the smaller distance and darker colors represent the larger distances between the data points. So, the lighter region contains a small group of tightly mapped data points.

In Fig. 6, input feature ApEn has lighter region than the other input feature. So, ApEn is the most compatible input feature mapping for the theta band among the 8 features. For Alpha and Beta bands, input feature scale variance has lighter region in lower corner of left side. But the other features have not has much lighter region than scale variance. So, scale variance is the compatible feature for the alpha and beta band among the 8 features. For more analysis, we plot the ApEn for theta band and scale variance for Alpha and Beta band separately which are shown in Fig. 7. This Figure concludes that right hand ApEn data distinguish the left hand data where right hand data in a level of mapping. Furthermore, scale variance of left and right hand data in Alpha and Beta band shows that right hand data is in upper level than the left the hand data. Figure 4 shows that much darker colors represent the input features roll off and MMAV. So, roll off and MMAV are the worst features of the Theta, Alpha and Beta band.

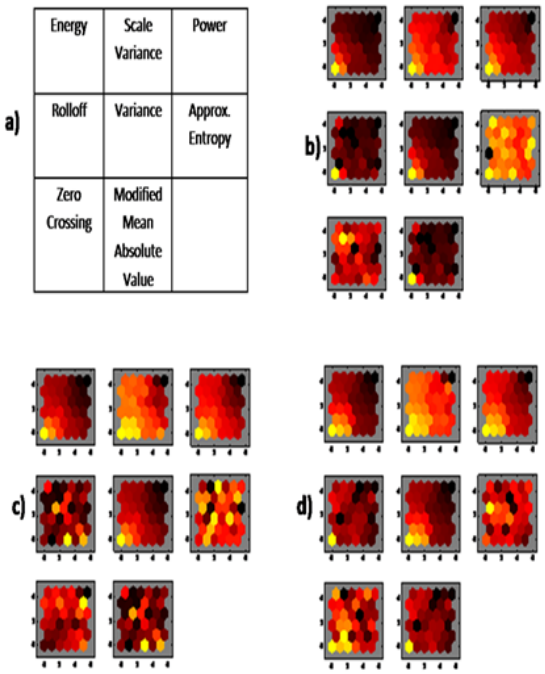


Fig. 6. Input Feature Mapping: a) Mapping identity of b), c) and d) figure, b) Mapping of Theta Band, c) Mapping of Alpha Band, and d) Mapping of Beta Band.

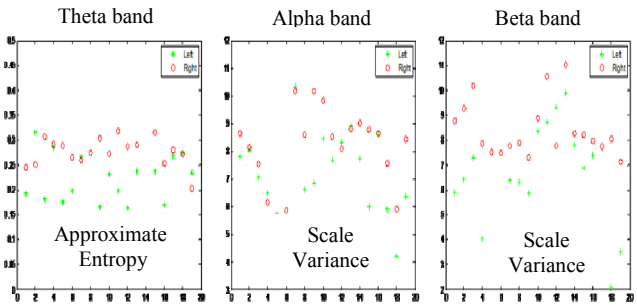


Fig. 7. Compatible features for different bands

IV. CONCLUSION

The objective of this work is to analyze the EEG signal for Left and Right hand movement. Alpha, Beta and Theta band of EEG signal are taken out through wavelet transform and extracted various features are trained into a multilayer back propagation algorithm to classify them. Classification rate shows that beta band (97.5%) has better convergence rate and mapping precision than the other two bands alpha (93.2%) and theta (87.8%). Another ANN Self-organizing feature maps (SOFM) is used to find the compatible feature for the individual band. SOFM shows that approximate entropy is the compatible feature for the theta band and for the alpha and beta band scale variance is the compatible feature for mapping. SOFM also shows that roll off and MMAV are the worst feature in alpha, beta and theta band for mapping. These findings might be useful in artificial hand movement through brain computer interfacing.

ACKNOWLEDGMENT

The authors would like to thank the authority of Khulna University of Engineering & Technology, KUET, and Higher Education Quality Enhancement Project (HEQEP, CP-2378), UGC, Bangladesh for their support.

REFERENCES

[1] M. A. Lebedev and M. A. L Nicolelis, "Brain-Machine Interfaces: Past, Present and Future," Trends in Neurosciences, vol. 29, no. 9, pp.536—546, 2006.

[2] L. R. Hochberg, M. D. Serruya, G. M. Friehs, J. A. Mukand, M. Saleh, A. H. Caplan, A. Branner, D. Chen, R. D. Penn, and J. P. Donoghue, "Neuronal Ensemble Control of Prosthetic Devices by a Human with Tetraplegia," Nature, vol. 442, no. 7099, pp. 164–171, 2006.

[3] J. R. Wolpaw, N. Birbaumer, W. J. Heetderks, D. J. McFarland, P. H. Peckham, G. Schalk, E. Donchin, L. A. Quatrano, C. J. Robinson, and T. M. Vaughan, "Brain-Computer Interface Technology: A Review of the First International Meeting," IEEE Transactions on Rehabilitation Engineering, vol. 8, pp.164—173, 2000.

[4] T. Ebrahimi, J.M. Vesin, and G.Garcia, "Brain-computer interface in Multimedia Communication," IEEE Signal Processing Magazine, vol.20, pp. 14-24, 2003.

[5] S. Sanei and J. Chambers, EEG Signal Processing, John Wiley & Sons Ltd, England, 2007.

[6] G. Pfurtscheller, C. Neuper, A. Schlögl, and K. Lugger, "Separability of EEG Signals Recorded During Right and Left Motor Imagery Using Adaptive Autoregressive Parameters," IEEE Transactions on Rehabilitation Engineering, vol. 6, no. 3, pp. 316-325, 1998.

[7] E. Haselsteiner and G. Pfurtscheller, "Using Time-Dependent Neural Networks for EEG Classification," IEEE Transactions on Rehabilitation Engineering, vol. 8, No. 4, pp. 457-463, 2000.

[8] D. Najumnissa and T. R. Rangaswamy, "Detection and Classification of Epileptic Seizures using Wavelet feature extraction and Adaptive Neuro-Fuzzy Inference System," International Journal of Computational Engineering Research, vol. 2, pp. 755-761, 2012.

[9] Y. Song, "A review of developments of EEG-based automatic medical support systems for epilepsy diagnosis and seizure detection," Journal of Biomedical Science and Engineering, vol. 4, pp. 788-796, 2011.

[10] I. Daubechies, "Ten lectures on wavelets," CBMS-NSF Regional Conference Series in Applied Mathematics, Society for Industrial and Applied Mathematics (SIAM), vol. 61, Philadelphia, PA, 1992.

[11] A. Subasi and E. Ercelebi, "Classification of EEG signals using neural network and logistic regression," Computer Methods and Programs in Biomedicine, pp.87—99, 2005.

[12] S. J. Husain and K. S. Rao, "Epileptic Seizures Classification from EEG Signals using Neural Networks," International Conference on Information and Network Technology, vol. 37, pp.269—273, 2012.

[13] S. G. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 11, pp. 674-692, 1989.

[14] M. S. Hussain, M. B. I. Reaz, F. Mohd-Yasin and M. I. Ibrahimy, "Electromyography signal analysis using wavelet transform and higher order statistics to determine muscle contraction," Expert Systems, vol. 26, pp.35-48, 2009.

[15] H. Ocak, "Automatic detection of epileptic seizures in EEG using discrete wavelet transform and approximate entropy," Expert Systems with Applications, Vol. 36, pp.2027–2036, 2009.

[16] S. Haykin, Neural Networks, a comprehensive foundation, Prentice Hall, 1999.

[17] L. Fausett, Fundamentals of Neural Networks, Prentice Hall, 1994.

[18] Available in: http://sites.google.com/site/projectbci/

A Formal Approach to Verify Software Scalability Requirements using Set Theory and Hoare Triple

Alim Ul Gias*, Mirza Rehenuma Tabassum*, Amit Seal Ami*, Asif Imran*, Mohammad Ibrahim*, Rayhanur Rahman* and Kazi Sakib†
Institute of Information Technology, University of Dhaka
Ramna, Dhaka–1000
*{bit0103, bit0129, bit0122, bit0119, bit0117, bit0101}@iit.du.ac.bd
†sakib@univdhaka.edu

***Abstract*—Scalability is the ability of a system to handle variation in execution environment and continuing to function in order to meet user needs. For ensuring scalability, it is important to verify that programmers are writing code that can scale. However, verifying scalability from code level has its own limitations as it did not receive adequate attention from researchers. This paper proposes a formal approach to verify scalability from the code level using set theory and Hoare triple. The method denotes variables and functions involving scalability through set notations. Hoare triple is used to measure the performance fulfillment with varying workload by a code segment given that certain code quality measure like caching or data compression is applied. The methodology is presented by means of an algorithm which strictly inhibits to passover a specific scalability requirement and requires to re-apply a quality measure until a specific requirement is being satisfied. The approach is applied for developing a real life online ticketing system and results show that it provides stable response time over a wide range of user requests. This indicates that the proposed approach is capable of ensuring scalability by verifying it from system's code.**

***Index Terms*—Software Scalability, Verification, Formal Method, Set Theory, Hoare Triple**

I. Introduction

Scalability is an expected requirement of a software system characterizing it's ability to fulfill quality goals like performance, throughput, etc. when characteristics of execution environment vary over a expected range [1]. It can be crucial for systems like e-commerce and social networking sites, web services and online ticketing systems which deal with growing number of users. These systems tend to experience unpredictable and extensively varying degrees of requests, especially during events such as breaking news, national occasions and Denial of Service (DoS) attacks [2].

Developing a scalable web application involves careful planning about coding so that it can make the application work properly under demanding workloads [3]. Therefore, it is essential to verify scalability from the code level for ensuring that the system can scale elastically when it is needed. Moreover, in case of some organizations inconsistent system performances due to lack of scalability can be fatal from the business perspective. Thus, during the development of those systems it is important to apply strictness to ensure none of the scalability requirements can remain unimplemented.

Although several works have been done on scalability framework and evaluation method [4], [5], [6], [7], [8], it is hard to find a significant work that specifically focuses on verifying scalability from the code level. State of the art works focus on architectural and design aspect [4], prediction of resources to make the system scalable [5] and scalability elicitation and analysis in requirement engineering phase [7], [6], [8]. However, it is arguable that existing works can be used to verify that whether the code can scale and impose the strictness regarding the implication of scalability.

This paper proposes a formal approach to verify the software scalability from the code level. The approach uses set theory [9] to define multiple sets of variables and their mapping functions involving system scalability. These sets include system's load extensive states, code quality measures like caching or data compression, performance expectation over a specific range of requests, state-quality compatibility pairs and code segments. Mapping functions include map from system states to code segments and compatibility pair to performance fulfillment with varying user requests. A performance requirement is fulfilled if its respective Hoare triple [10] including code quality measure, code segment and performance expectation holds. The whole procedure is represented by an algorithm that strictly inhibits to leave a specific requirement without being implemented and requires to re-apply certain quality measure until the expected performance goal is being achieved.

The approach was applied to develop the prototype of an online ticketing system. Before the implementation, requirement analysis was performed to identify load extensive states, code quality measures and expected performance goals with respect to specific number of user requests. The prototype is tested using a software named *ab*[1] which is a HTTP server benchmarking tool provided by Apache. The software was used to generate a large amount of user traffic varying the number of requests at a time from 100 to 1000 for a total period of 10 second keeping the concurrency level to 100. Results show that the system is responding within the expected response time of 15 ms and 30 ms with minor fluctuations due to multiple web server processes and caching limitations.

[1] httpd.apache.org/docs/2.2/programs/ab.html

These results assure that the proposed methodology is capable of ensuring scalability from the code.

The rest of the paper is organized as follows: Different frameworks for scalability analysis and elicitation is discussed in Section II. The proposed approach is defined in Section III. Section IV presents the case study of an online ticketing system on which the proposed approach is being applied. Lastly, this paper is concluded with a discussion about few observations visualized from results of the case study and future research directions.

II. Literature Review

Scalability is a major concern for software development but not much work has been done on it. Earlier, scalability as a requirement was overlooked while developing software systems for lack of techniques for eliciting and testing scalability requirements [7]. Most of the works focused on building framework for scalability. Some of the significant works is going to be discussed briefly in this section.

Scalability was analyzed for Admission Control Server Based Resource Allocation (ACS-RA) in [4]. It was then compared with Resource Reservation Protocol (RSVP) [11], [12] approach. The work was done by analyzing a framework for reserving resources according to the RSVP model in a network cloud. They analyzed the scalability in the worst case scenario which is a single centralized Admission Control Server for a given differentiated services cloud [13], [14]. They showed that ACS-RA is better than RSVP approach in the aspect of scalability. However, investigation on the comparison of the user plane performances was expected in their work. Assuming a single centralized ACS to have a worst case scenario proved their claim that ACS-RA architecture scales better than RSVP successfully but some other scenarios should be added.

Lloyd G. Williams et al. presented Quantitative Scalability Evaluation Method (QSEM) to directly quantify the scalability of a software system [5]. Their method identifies specific quantity of resources needed to be used for ensuring scalability. They have shown results from a case study after following seven steps of QSEM. It uses straightforward measurements of maximum throughput at different amount of processors or nodes of the targeted system. Identifying critical use cases, selecting representative scalability scenarios, determining scalability requirements, planning measurement experiments, performing measurements and presenting the result are the steps of QSEM to complete the scalability analysis. The steps of their method shows how to evaluate scalability from the very first phase with identifying critical use cases. Their evolution method can measure scalability requirement quantitatively which can surely help to figure out the actual requirement and outcome. This approach is good to verify scalability from use case to application level but a code verification step could make this approach more attractive.

As scalability is not defined and understood clearly, a framework for accurately characterizing and analyzing scalability requirements is presented in [6]. Resolving intuitions, ambiguities and inconsistencies of scalability is the primary goal of that framework. Authors claimed that their framework precisely detects dependency relationships of scalability and describe important steps of scalability analysis. The framework treats scalability as a multi-criteria optimization problem. They used principles of microeconomics to support their analysis.

Scalability has not been given the systematic treatment it deserved claimed in [7]. A framework for characterization and analysis of software system scalability for addressing the mentioned issue has been proposed. Application of the popular method goal-oriented requirements engineering (GORE) [15], [16] is described with a case study of a real word software system in the paper. The study found appropriateness of GORE as well as it's limitations which are shown in their result. They claimed that GORE lacks techniques to elaborate goal models with respect to scalability goals. Though the advantages of using GORE are discussed quite elaborately, drawbacks are not discussed properly.

Leticia Duboc et al. worked for elaborating specific scalability requirements through identifying goal-obstacles to satisfy those requirements [8]. Their method systematically elaborate and analyze scalability requirements. Result of their research was presented from a case study in which the method was applied to a complex, large-scale financial fraud detection system. Method of this research was formulated from KAOS [17], [18] goal-oriented requirements engineering method. They used KAOS for the elaboration of goal models, management of conflicts between goals and selecting alternative system designs. The key issues they identified in their work can help to fix the developers view towards scalability requirements.

The review of state of the art works shows that though multiple issues have been considered in proposing a method for scalability prediction, elicitation and analysis, none of the work offer an approach to verify scalability from the code level. However, in order to have consistent system performance over varying workloads it is essential to write scalable code and a formal approach should be provided for its verification. This paper introduces one such approach for verifying scalability using set theory and Hoare triple and strictly requires to fulfill all scalability requirements.

III. Formalizing Scalability Verification

We define a set of states which will receive excessive load as $\mathcal{S} = \{s_1, s_2, ..., s_n\}$. Let $\mathcal{C} = \{c_1, c_2, ..., c_n\}$ denote the set of code segments that eventually forms the whole system. We also define a mapping function $\psi : \mathcal{S} \rightarrow \mathcal{C}$ that will relate a specific load extensive state to a code segment. The set of certain code quality measures that should be applied for allowing the code to scale is defined as $\mathcal{Q} = \{q_1, q_2, ..., q_m\}$.

It is obvious that all quality measures will not be applicable to a specific state. Thus we define a set of compatibility pairs where each pair include a specific state and quality measure. This is denoted as $\Omega = \{\omega_1, \omega_2, ..., \omega_p\}$ where $\omega_t = \{s_i, q_j\}, i \leq n$ and $j \leq m$. We define the set of expected performance outputs with varying workloads as

$\mathcal{K} = \{k_1, k_2, ..., k_p\}$ where k_t is a output of the expected function ξ with respect to ω_t which means $k_t = \xi(\omega_t)$.

In order to fulfill a specific scalability requirement, each compatibility pair has to satisfy the function $\phi : \Omega \rightarrow \{0, 1\}$. The output of the function is defined as follows:

$$\phi(s_i, q_j) = \begin{cases} 1 & \text{if the Hoare triple holds} \\ 0 & \text{otherwise} \end{cases}$$

A Hoare triple of the form $\{\mathcal{Q}\}\ \mathcal{C}\ \{\mathcal{K}\}$ will hold if after executing $c_i \in \mathcal{C}$ derived from $s_i \in \omega_t$ using ψ, the expected performance output $k_t = \xi(\omega_t)$ is satisfied given that $q_j \in \omega_t$ is implied.

Algorithm 1 illustrates certain steps that need to be performed for verifying the system scalability using the proposed formal approach. The algorithm requires the scalability requirement specification as an input and it will ensure that all requirements are satisfied. Initially load extensive states $\mathcal{S}$, code quality measures $\mathcal{Q}$, compatibility pairs Ω and expected outcomes $\mathcal{K}$ will be defined. During the implementation phase, $\mathcal{C}$ and mapping function ψ will be defined. After finishing the implementation it has to be ensured that all Hoare triple hold. If a Hoare triple does not hold then the code chunk must be redefined with respect to code quality measures until the triple holds.

Algorithm 1 Scalability Verification

Require: Scalability requirement specification
Ensure: All the scalability requirements are satisfied
Begin
Identify load extensive states $\mathcal{S}$
Define code quality measures $\mathcal{Q}$ and compatibility pairs Ω
Define expected outcomes $\mathcal{K}$
As the system being implemented define $\mathcal{C}$ and mapping function ψ
while $|\Omega| \neq 0$ **do**
    Use ψ to get c_i from s_i
    if $\{q_j\}\ c_i\ \{k_t\}$ holds **then**
        $\Omega \leftarrow \Omega \setminus \{\omega_t\}$
    else
        redefine c_i with respect to q_j
    end if
end while
State that all the scalability requirements are satisfied
End

IV. Case Study

A case study was conducted on an online ticketing system to illustrate how the proposed method can be used to verify system scalability. This section discusses the details of that case study including the system scenario, implication of the proposed approach on developing the prototype of the ticketing system and results obtained after load testing the prototype for measuring system consistency.

A. System Scenario

The user will initially have a view of the tournament fixture for the current season. The fixture will show the match number, teams, venues, date and time of matches. An option will be provided beside the matches yet to occur, enabling the user to buy ticket. After choosing a match to buy ticket, the user will be given the option to choose the gallery, block number, row and the seat number. After choosing the options the price of the ticket (according to the option) will be shown. If the user selects the ticket to buy then s/he will be considered as a buyer and asked for providing necessary information which are name, address and mobile phone number.

An SMS having a secret code will be send to the submitted mobile number. The buyer will then be provided the option to submit that code and if the code is correct the submitted information will be stored in the database. A unique serial number will be generated to keep track of the buyer and s/he will be taken to complete the payment. After receiving the confirmation that the money has been paid, a receipt in the PDF format, having the generated serial will be provided by the system. The buyer can use that receipt to collect his/her ticket from the nearest ticket collection center.

B. Exercising the proposed approach

After analyzing the whole process of this online ticketing system using the approach of [5], some of the states are considered as critical which must be scalable. It is showed by the system scenario that several requests from users and options providing by systems can be programmed in such a way which can ensure the expected performance. Data compression and caching are the two techniques implemented here for scalability. Only two techniques are used to keep it simple for this case. Code sections of critical states have been fetched where cache and compression are applied.

Scalability is verified by the algorithm showed in earlier section with Hoare Triple and Set Theory. Code sections with the selected states are mapped and techniques are implemented upon them. Different states have different quality measures. Data compression and caching are the two quality measures where related states are coupled to form the compatibility pair. Expected outcomes for the compatibility pairs are made from the user statistics and surveys as discussed in [5].

Critical states are first taken and related quality measures are mapped. Viewing fixture by the user, gallery selection, ticket buying option provided by the system, adding user information and ticket purchasing information of users, updating ticket availability and generating receipt are the critical states of online ticketing system. The states which affects the systems scalability is referred as critical states. Two of the critical states are shown here as representing states.

Fixture viewing state can be cached to ensure scalability for varying execution environment. The assumptions were based on statistics that 10,000 users may request to view fixture per seconds with a 56 Kbps connection. Test was done in Local area network which indicates that the requests should be served within 15 ms. When the state is s_f and the quality measure cache is q_c, it makes the compatibility pair ω_m. Expected outcome k_m should be within 15 ms. $\{q_c\}\ c_f\ \{k_m\}$ hold, if the outcome after test is less than 15 ms.

Generating receipt is another critical state where a receipt of the purchase is provided to the user. Compression is applied

TABLE I: Formal Scalability Verification of View Fixture and Generate Receipt Modules

Name of the state	View Fixture (s_f)	Generate Receipt (s_r)
Quality Measure ($\mathcal{Q}$)	Cache (q_c)	Compression (q_p)
Code Chunk ($\mathcal{S} \rightarrow \mathcal{C}$)	(c_f)	(c_r)
Expected Outcome ($\mathcal{K}$)	Response time within 15 ms up to 10,000 (k_m)	Response time within 30 ms up to 10,000 (k_n)
Hoare Triple	$\{q_c\}\ c_f\ k_m$	$\{q_p\}\ c_r\ \{k_n\}$

in this state to compress the receipt and pass over the network. If a 311 KB receipt file becomes 259 KB after compression, then it should take around 30 ms per request which was stated in the requirements specification document. The compatibility pair ω_n is made with the state s_r and q_p. $\{q_p\}\ c_r\ \{k_n\}$ hold, if expected outcome k_n is less than 30 ms. Table I summarizes the formal verification of mentioned two states.

C. Results

The prototype of the system was developed using PHP and MySQL Database on 32 bit Linux Mint 15 machine with Intel Core-i3 processor, 4GB RAM. It was hosted using Apache server and load tested using the benchmarking tool *ab* provided by Apache. Two modules - *View Fixture* and *Generate Receipt* were implemented each having two versions. In case of version one for the View Fixture module, caching was applied following the proposed approach and the other version involved no caching. Similarly, in case of Generate Receipt module a version included data compression and the other did not. User requests were generated through *ab* by varying the number of requests at a time from 100 to 1000 for a total period of 10 second keeping the concurrency level to 100. The response times in each of these four cases are reported in Table II and III.

The graph of Table II and III is shown in Figure 1 and 2 respectively. From the graph it is observed that the response time in case of caching is much more less than caching being not applied. However, in case of compression though the same thing is suppose to happen, the opposite phenomenon is observed. This is detailed discussed in Section V

V. Discussion

An important observation can be found from Figure 1a. Here, it is observed that even though the time required to serve request remains somewhat consistent, there are several significant spikes in terms of time required. This phenomenon is observed because when a request is submitted at first, it is processed to be stored in cache for future usage. As a result, time required for handling further requests within the same axes are shown to be smaller comparatively. This can be seen in the Figure 1a where number of requests was between 200-400 range, where initial time required was from 13.5 to 14.5 unit times, whereas further requests required around 13 unit times only. The other spikes observed, particularly within 600-800 number of requests range and 800-1000 number of requests range however, are not observed in cases of initial time required. This happened because the system, hosting the web server, was tasked for serving multiple types of requests at the same time and processed other types of data similar to any real application server. As a result, random spikes occurred throughout the observation. However, overall time required per requests remained consistent.

In Figure 1b, performance in terms of no cache is depicted. Here, the observation is that time required to serve requests was high right from the start. The peaks are observed at specific number of requests, such as 400, 600 and 800. The reason is that because of lack of cache, the time require to process requests was always high. The decreased value is observed because the number of requests was not made continuously, but discretely. To explain, values taken started from 200 requests, with increment of 200 requests per round, to 1000 requests. It is also observed that with the number of requests handled, the time required to handle requests increased as well, making this approach inappropriate for scaling.

In Figure 2a and 2b, the performance is shown in terms of time required per request for compressed and non compressed data. However, unlike known observations, it is observed that performance of compressed data is not good compared to uncompressed data. The reason is that the experiment was conducted in Local Area Network, where there is low latency for communication between client and server end. As a result, the time required to compress data for each request became significant. There are also several spikes observed randomly, which occurred due to the real system like property of our environment serving multiple types of request simultaneously instead of just allocating processor for providing web services.

VI. Conclusion and Future Work

This paper proposes a formal approach that utilizes Set theory and Hoare triple to ensure proper exercise of each scalability requirements. The importance of verifying scalability from the code level was highlighted as it can make applications consistent with respect to performance goals despite rapid increase of user requests. The proposed method applies a strictness that none of the scalability requirements can be skipped and allows re-applying code quality measures, like data compression and caching, until fulfilling certain performance goals.

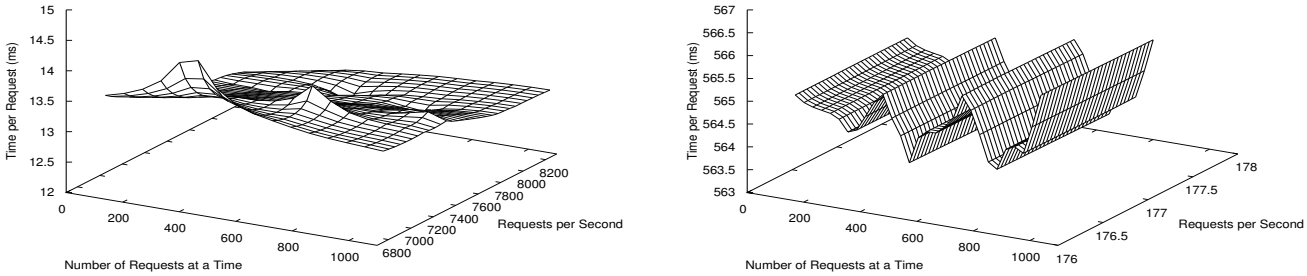
For experimental purpose, an online ticketing system was developed using traditional approach without following the proposed method. Requests varying within a range of 100-1000 with constant concurrency of value 100 were sent to the application for a time period of 10 second. Another application was developed that followed the proposed approach

TABLE II: Performance Comparison of Two Versions of the Generate Receipt Module

Number of Req. at a Time	Caching Applied		No Caching Applied	
	Req. per Second	Time per Req. (ms)	Req. per Second	Time per Req. (ms)
100	8282.82	12.073	176.98	565.027
200	7777.96	12.857	176.76	564.783
300	7768.76	12.872	177.29	564.051
400	6797.78	14.711	176.75	565.764
500	7976.41	12.537	177.39	563.743
600	7699.57	12.988	177.09	564.676
700	7158.67	13.969	176.64	566.113
800	7806.78	12.809	177.47	563.468
900	7481.52	13.366	177.19	564.37
1000	7351.29	13.603	176.22	567.468

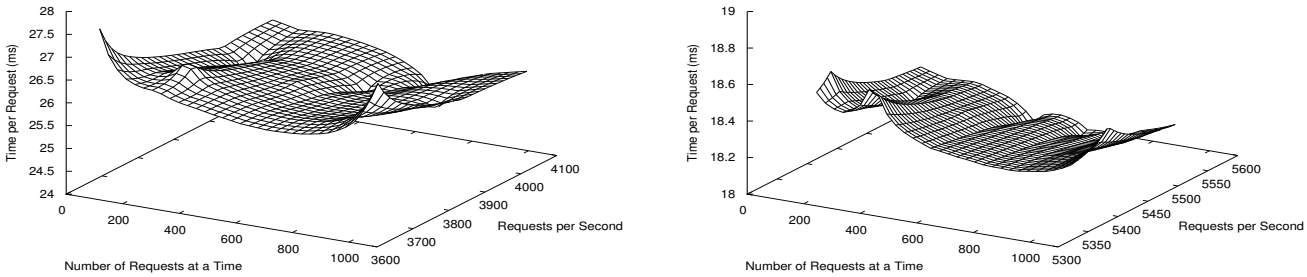
TABLE III: Performance Comparison of Two Versions of the Generate Receipt Module

Number of Req. at a Time	Compression Applied		No Compression Applied	
	Req. per Second	Time per Req. (ms)	Req. per Second	Time per Req. (ms)
100	3614.46	27.667	5392.12	18.546
200	3955.13	25.284	5485.54	18.23
300	3704.21	26.996	5368.53	18.627
400	3654.21	26.312	5469.02	18.285
500	3841.4	26.032	5533.1	18.073
600	3957.48	25.269	5493.3	18.204
700	4095.89	24.415	5531.74	18.078
800	3890.11	25.706	5542.59	18.042
900	3917.67	25.525	5501.72	18.176
1000	3679.44	27.178	5430.53	18.414



(a) Caching Applied (b) No Caching Applied

Fig. 1: Performance visualization in terms of caching being applied and not applied



(a) Compression Applied (b) No Compression Applied

Fig. 2: Performance visualization in terms of compression being applied and not applied

and similar number of requests were sent. The response time was recorded for each of those cases and results show that the later system significantly outperforms the first one. These results illustrate that the proposed approach can be used to ensure scalability of a system from it's code.

The proposed method considers fulfilling each scalability requirement as a binary function and functions specific to all requirements must return true for a successful exercise of the approach. A further research challenge could be applying specific weight to each scalability requirement and consider the approach as a Fuzzy system. Work has to be done on how to assign those weight and obtaining the thresh-hold value based on which the success of the approach will be measured.

Acknowledgement

This research has been supported by the University Grant Commission, Bangladesh under the Dhaka University Teachers Research Grant No-Regi/Admn-3/2012-2013/13190.

We want to express our gratitude to Mr. Shah Mostafa Khaled, Assistant Professor, Institute of Information Technology, University of Dhaka for helping us in modeling the proposed method.

References

[1] D. S. Rosenblum, "Software system scalability: concepts and techniques," in *Proceedings of the 2nd India software engineering conference*. ACM, 2009, pp. 1–2.

[2] C. Olston, A. Manjhi, C. Garrod, A. Ailamaki, B. M. Maggs, and T. C. Mowry, "A scalability service for dynamic web applications," in *Proceedings of the 2nd Biennial Conference on Innovative Data Systems Research*, 2005, pp. 56–69.

[3] A. Otto, "Write scalable code," *Code Project [Online]*, 2011.

[4] A. Detti, M. Listanti, S. Salsano, and L. Veltri, "Supporting rsvp in a differentiated service domain: an architectural framework and a scalability analysis," in *International Conference on Communications*, vol. 1. IEEE, 1999, pp. 204–210.

[5] L. G. Williams and C. U. Smith, "Qsemsm: Quantitative scalability evaluation method," in *Proc. CMG*, 2005.

[6] L. Duboc, D. Rosenblum, and T. Wicks, "A framework for characterization and analysis of software system scalability," in *Proceedings of the the 6th joint meeting of the European software engineering conference and the ACM SIGSOFT symposium on The foundations of software engineering*. ACM, 2007, pp. 375–384.

[7] L. Duboc, E. Letier, D. S. Rosenblum, and T. Wicks, "A case study in eliciting scalability requirements," in *Proceedings of 16th International Requirements Engineering*. IEEE, 2008, pp. 247–252.

[8] L. Duboc, E. Leiter, and D. S. Rosenblum, "Systematic elaboration of scalability requirements through goal-obstacle analysis," *IEEE Transactions on Software Engineering*, vol. 39, no. 1, pp. 119–140, 2013.

[9] B. Potter, D. Till, and J. Sinclair, *An introduction to formal specification and Z*. Prentice Hall PTR, 1996.

[10] C. Stirling, "A generalization of owicki-gries's hoare logic for a concurrent while language," *Theoretical Computer Science*, vol. 58, no. 1, pp. 347–359, 1988.

[11] L. Zhang, S. Deering, D. Estrin, S. Shenker, and D. Zappala, "Rsvp: A new resource reservation protocol," *IEEE Network*, vol. 7, no. 5, pp. 8–18, 1993.

[12] G. R. Mallofre, "Resource reservation protocol (rsvp)," in *Seminar on Transport of Multimedia Streams in Wireless Internet, Department of Computer Science University of Helsinki, Finland*. Citeseer, 2003.

[13] E. Nyberg, S. Aalto, and R. Susitaival, "A simulation study on the relation of diff-serv packet level mechanisms and flow level qos requirements," in *International Seminar on Telecommunication Networks and Teletraffic Theory*, 2002.

[14] R. Balmer, F. Baumgarter, T. Braun, and M. Gunter, "A concept for rsvp over diffserv," in *Proceedings of Ninth International Conference on Computer Communications and Networks*. IEEE, 2000, pp. 412–417.

[15] A. van Lamsweerde, "Goal-oriented requirements enginering: a roundtrip from research to practice [enginering read engineering]," in *Proceedings of 12th International Requirements Engineering Conference*. IEEE, 2004, pp. 4–7.

[16] A. Lapouchnian, "Goal-oriented requirements engineering: An overview of the current research," *Technical Report http://www.cs.toronto.edu/~alexei/pub/Lapouchnian-Depth.pdf, University of Toronto*, 2005.

[17] R. Darimont, E. Delor, P. Massonet, and A. van Lamsweerde, "Grail/kaos: an environment for goal-driven requirements engineering," in *Proceedings of the 19th International Conference on Software engineering*. ACM, 1997, pp. 612–613.

[18] W. Heaven and A. Finkelstein, "Uml profile to support requirements engineering with kaos," *IEE Proceedings-Software*, vol. 151, no. 1, pp. 10–27, 2004.

A Helper Initiated Distributed Cooperative Medium Access Control (MAC) Protocol for Wireless Networks

Md. Ruhul Amin[1], Shafika Showkat Moni[2], Shamim Ara Shawkat[1] and Mohammad Shah Alam[1]

[1]Institute of Information and Communication Technology
Bangladesh University of Engineering and Technology
Dhaka, 1000, Bangladesh
Email: ruhulamin@live.com

[2]Department of Computer Science and Engineering
Rajshahi University of Engineering and Technology
Rajshahi, Bangladesh

***Abstract*—Spatial diversity for wireless transmission requires more than one antenna at the transmitter. However, mobile devices are usually limited by size, so installation of multiple antennas increases the hardware complexity significantly. Due to the omnidirectional nature of wireless signal, a data transmission between a source node and a destination node can be overheard by many other neighbor nodes. By exploiting this characteristic, a number of recent research activities on cooperative Medium Access Control (MAC) have been devised where low data rate stations are assisted by the high data rate stations in forwarding data traffics. Therefore, wireless devices with a single antenna can effectively form a virtual array of antennas by sharing each other's' antennas in a multiuser environment. To facilitate cooperative communication in the data link layer requires a great deal of attention. In designing a cooperative MAC, selection of relay stations is the most important criterion. In this paper, we propose a distributed cooperative MAC protocol where a potential relay node initiates itself to participate in the cooperation by calculating supported data transmission rate between source to relay and relay to destination links. A mathematical analysis of our proposed scheme is derived and throughput of the proposed scheme is then compared with that of the existing IEEE 802.11 DCF MAC. Numerical results show that our proposed scheme can increase the throughput of any IEEE 802.11 low data rate station comprehensively.**

Keywords—Cooperative Networking, IEEE 802.11 DCF, MAC,Helper Initiated Medium Access Control (MAC) Protocol, RTS, CTS,CoopMAC.

I. INTRODUCTION

IEEE 801.11 DCF MAC is the de facto standard for wireless LANs' medium access control. Multi rate capabilities of IEEE 802.11 wireless LANs' standards have encouraged the researchers to divert the cooperative communication from physical layers perspectives to MAC layer. Physical layer cooperative communication schemes have facilitated the surrounding wireless stations of both source and destination to process and transmit the overheard signal on behalf of the source. Due to fading and interferences induced into the transmitting signals limits data rates supportable between source and destination. Researchers have not confined themselves only into the physical layer for cooperative communication but have extended their works into the data link layer. In wireless LANs i.e. IEEE 802.11b, the presence of more 1 Mbps stations significantly decreases the system throughput. IEEE 802.11 DCF MAC provides equal access in time to all of the contending stations for data transmission. Slower stations consume more time to transmit data that consequently decreases the probability of accessing the medium for other stations. Therefore, transmission capabilities of other higher data rate stations are wasted. Thus, system throughput decreases significantly. To mitigate this drawback new technologies such as multiple input multiple output (MIMO) has been developed to increase transmission rate by increasing number of bits per hertz.

Now-a-days, the size of wireless terminal and hand held devices have been decreased. So, multiple antennas installation in a single small wireless device is highly complex. Moreover, architectures of wireless devices may not congruent with the minimum physical layer requirements of MIMO. Cooperative communication exploits the broadcast nature of wireless transmission. While a wireless station transmits, that transmission is heard by all other stations within the transmitter proximity. Focus of Cooperative MAC is to find out an appropriate relay to help out the sender and receiver pair in transmitting their data frames faster and with reliability. In Cooperative MAC, a transmitter sends its data to a helper and the helper forwards the data packet to the receiver. Thus the data frame is sent to the receiver by two hops instead of single hop. Cooperative MAC helps in the formation of virtual antenna arrays through individual terminals distributed in space. Thus, virtual antenna array requires a significant amount of coordination. The lesser the coordination overhead incurs in cooperation the more

throughput efficiency could be achieved. This spatial diversity without equipping individual nodes with multiple antennas provides reliable data transmission. Two popular mechanisms in physical layer cooperation are decoded-and–forward and compress and forward [1]-[2]. According to the former, helper stations decode and re-encode the received signal before forwarding it to the destination. In the later technique, the helper first compresses the received signal before forwarding it to the destination.

In cooperative MAC design, selection of a proper helper is the most significant issue. An appropriate helper selection strategy can reduce the overhead in cooperative communication significantly. A bad helper selection strategy in worst case may abruptly reduce throughput performance of the communication system.Initiating a cooperative mode should be selective. Sometimes a sender station may not need to take help from the relay. If it initiates cooperative transmission when it does not need to do so will badly affect its throughput gain. In this paper, we have proposed a helper initiated cooperative MAC. In helper initiated cooperative technique, relay station will determine whether initiating cooperation for the ensuing data transmission shall be beneficial. The relay selection in our proposed protocol is distributed. Unlike other works [4], in our paper we have shown how the relay can be the initiator of the cooperative communication in data link layer. We have also derived equations for the helper initiated distributed cooperative MAC and analyze the throughput performance of our proposed technique. The rest of the paper is organized as follows: Section II gives the related works of cooperative MAC design; Section III describes the legacy IEEE 802.11 DCF MAC in brief. In Section IV we describe our proposed helper initiated distributed cooperative MAC, and section V describes the throughput analysis and numerical results. In Section VI concludes this paper.

II. RELATED WORKS

Research communities are continuously devising new methods to employ spatial diversity in single antenna based nodes wireless networks. Several physical layer methods have been studied. Among them, amplify and forward and decode and forward are the two well-known methods [1]-[2]. Only a few research studies have focused on cooperative MAC protocol design. rDCF [3](Relay Enable DCF) has exploited the multi rate nature of IEEE 802.11. In this work a helper is selected taking the channel condition into consideration. A helper is selected using a triangular hand shake among sender, helper and receiver in predetermined manner. Pei Liu has designed a cooperative MAC [4]. In this works a helper is selected by the sender in proactive manner. This technique lacks the addressing of helper selection in distributed fashion within the sender and receiver vicinity. Another Cooperative MAC protocol for Vehicular Networks is proposed in [5]. This works has addressed the issue of mobility but the best helper selection takes higher overhead, which degrades the system throughput. Hangguan Shan and others have proposed a cross layer cooperative MAC protocol [5]. A helper is selected from a predetermined cooperative region and the collision probability in selecting a single helper is very high. In [6], Xin has designed a Cooperative MAC for multi hop wireless networks in which helper selection is initiated by the receiver only when a data packet is not received correctly. Data transmission from the helper is prone to collision. S. Bharati [7] has proposed a cooperative Ah hoc MAC for Vehicular Networks. In his work, he has applied a time division multiple access schemes in selecting helper which incurs the disadvantages of TDMA and also has modified the legacy IEEE 802.11 DCF Mac frame format. Other cooperative MAC protocols have also been proposed in [8] [9]. All of them are proactive helper selection techniques. Those works are not backward compatible with IEEE 802.11 Mac. Therefore, devising an appropriate relay station selection strategy in cooperative MAC design for wireless Networks is not straightforward.

III. IEEE 802.11 DCF MAC

DCF (Distributed coordination function) is based on carrier sense multiple access with collision avoidance (CSMA/CA) protocol. In IEEE 802.11 DCF, shared medium is dedicatedly allocated to the contending stations assigned to two communicating stations. While the medium is assigned to a communication party other stations inside their proximity cannot access or transfer data during that data transmission session. There are two access mechanisms in IEEE 802.11 DCF MAC. Its frame structures are shown in Fig. 1.

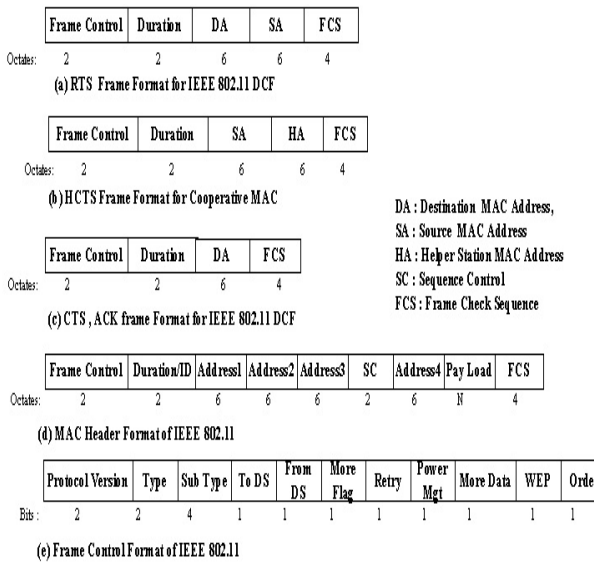


Fig. 1. Frame Structure of IEEE 802.11 DCF MAC

A. Basic Access Mechanism

In order to mitigate the collisions within a collision domain, every wireless station using IEEE 802.11 DCF follows random binary exponential back off algorithm (BEB). This randomized algorithm controls wireless stations in accessing the wireless medium. Particularly, the time is slotted. Every station is allowed to transmit at beginning of a time slot. A slot time is the time required to detect the transmission of a frame from any other station. A station cannot transmit until its back off counter is decremented to

zero. The back off counter of a station is decremented while the medium is sensed idle and frozen while it is busy. In order to track back off counter, every station maintains a contention window (CW). Each station has a maximum retry count which indicates the number of times a frame can be retransmitted due to its transmission failure. The value of CW is chosen between (0, CW-1). Value of CW depends on the number of failure of a frame transmission. At the first time, CW is set to its minimum value, CW_{min}. If a collision during a frame transmission, CW value is doubled. CW has a maximum value which is $CW_{max}=2^k x CW_{min}$, where k is the maximum number of retry limit. If the CW reaches to its maximum, it will remain in its maximum until the CW is reset to CW_{min}. CW is set to its minimum in case of a successful frame transmission or if the maximum retry limit for a frame is reached.. Basic access process is shown in Fig. 2.

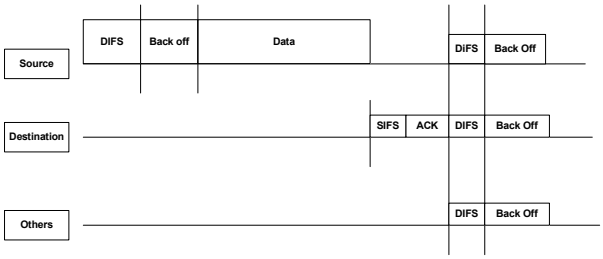


Fig.2. IEEE 802.11 Basic Access Mechanism

If a station has a frame, it senses the medium for DIFS (Distributed Inter frame Space) time. If it finds the medium idle during that time and its back off counter is zero, it simply sends the frame. After frame is received correctly in the receiver, the destination station sends an acknowledgement (ACK) of the receipt frame to the sender. The destination station sends ACK after SIFS (Short Intra frame Space time. The length of SIFS is shorter than that of DIFS. In IEEE 802.11 DCF basic mechanism, a collision may occur during the Frame transmission and ACK transmission. In both cases the sending station retransmits the frame.

B. IEEE 802.11 RTS-CTS Access Mechanism

Due to the broadcast nature of wireless signal transmission, basic access mechanism have two problems i.e. hidden station problem and exposed station problem. To deal with the former problem, IEEE 802.11 DCF has employed RTS and CTS control frames.

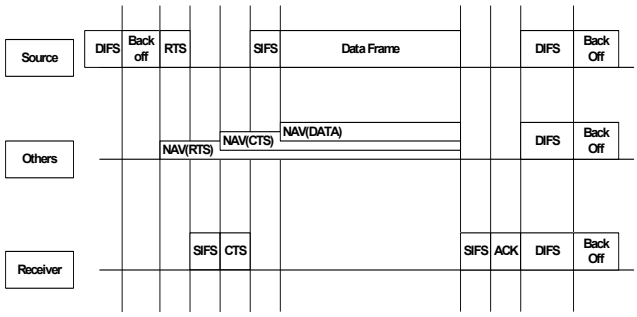


Fig. 3 IEEE 802.11 DCF RTS-CTS Access

Introduction of these two control frames minimizes the extra time required due to more frequent collisions in heavily loaded wireless transmission system. Frame structures of RTS-CTS access mechanism are shown in fig.3.

If a station finds the medium idle for DIFS time, it sends a RTS frame to the destination station. If the RTS frame is received by the destination correctly, it replies to the source station with a CTS frame after a SIFS time interval. The source after upon receiving a CTS frame from the receiver, it sends the frame spending a SIFS time interval. After the data frame is received by the destination station, an ACK frame is sent to source after SIFS time. The successful transmission of RTS and CTS frames between the two communicating stations keeps the medium reserved for a particular data transmission session. Therefore, the collision time is kept minimized for a long data frame which was imminent in IEEE 802.11 DCF basic access mechanism. The sending station uses a parameter RTS threshold in determining which mechanism it should be invoked. RTS threshold is determined by the size of the payload the frame will be carrying. In this mechanism, a collision can only be occurred during RTS frame transmission.

IV. PROPOSED COOPERATIVE DISTRIBUTED MAC

A. Problem Formulation

In IEEE 802.11 DCF, all of the wireless stations within a single collision region get equal medium access facility in the long run. It means in stable condition system allows all the wireless stations equal amount of time for data transmission. This phenomenon has motivated the researchers in designing of a Cooperative MAC. A number of low and high transmitting stations usually reside in a communication system. Obviously, low transmitting stations take much time to transmit their data comparing to the high transmitting stations. Due to those low transmitting stations, high transmitting station get much lower time to access the medium. Therefore, it leaves the system throughput to be degraded. So, if the low transmitting stations are aided by the others high rate stations in relaying their data to the destination, system performance shall be increased. This is because of the fact that lower transmitting stations will consume less time giving the others transmitting stations more access time for their data transmissions.

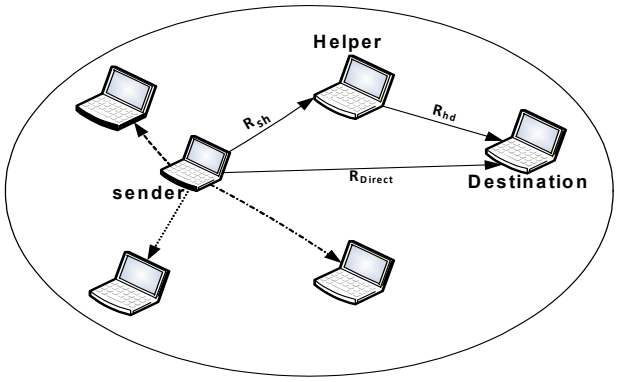


Fig. 4. Coooperatoin Scenerio

Multi rate capabilities of the wireless stations have facilitated the research community to ponder in designing cooperative MAC. Moreover, cooperative MAC may focuses on virtual MIMO by using single antenna on each wireless station exploiting the spatial diversity in fig 4.

B. Cooperation Initiation

Each wireless station shall act as a sender or receiver and as well as a relay. In IEEE 802.11 MAC if a data packet is received by a station that is not its intended destination of the packet, the station shall discard the packet by setting its NAV (Network Allocation Vector) to the value of the duration field tagged inside the received packet header. This broadcast nature of the wireless transmission facilitates in designing our proposed Cooperative MAC. Every wireless node within the transmission region of both the sender and receiver can hear the maximum transmission rate intended to be used between the communicating pair. Those intermediate stations may act as the helper stations which could accelerate ongoing data transmission rate.

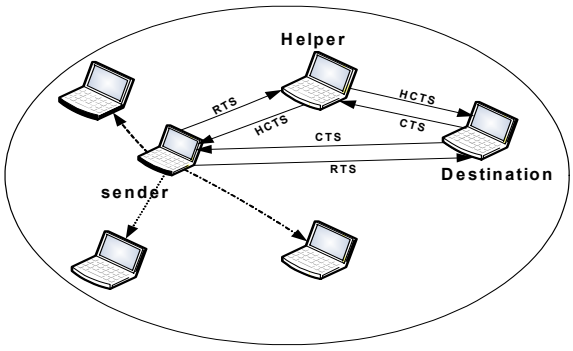


Fig.5. Control Frame Handshaking in Cooperative MAC

C. Relay Station Selection

In this proposed protocol, each wireless station shall maintain a CoopTable like in [4]. In IEEE 802.11, the control frames and headers are always modulated at the base rate .i.e. 6Mbps for IEEE 802.11a and 1 Mbps for IEEE 802.11b wireless networks [3]. Unlike [4] a *Coop Table* in this proposed protocol has the format like in Table 1 and each row in the table keeps record for only one potential relay stations. Each wireless node in this phase calculates the received overheard RTS and CTS signal strength (RSS) from every other station in its proximity. Station which receives both RTS and CTS control frame from sender and receiver respectively assumes itself as a potential relay station for the ensuing data transmission session. Each relay station calculates maximum direct data rate between itself-sender and itself –receiver links through measuring the signal strength of RTS and CTS respectively. In either case, if maximum transmission rate is greater than a threshold (5.5 Mbps for IEEE 802.11 b), it adds the MAC address of sender of RTS/CTS into its CoopTable. Figure 5 depicts the handshaking mechanism of RTS-CTS-HCTS control frames in our proposed cooperative MAC. It also adds Timestamp of the last received packet into its CoopTable for each received RTS/CTS frames. If a relay station already has an entry of a neighbor station in its CoopTable from which it has just received an RTS/CTS and supported direct transmission rate is below the threshold, it simply deletes entry of that neighbor station. This filtering ensures that only the effective data transmission would take place through this helper and as well as keeps the size of the CoopTable minimum. A station considers itself as an effective relay if it finds that

$$\left(\frac{1}{R_{sh}} + \frac{1}{R_{hd}}\right)^{-1} > R_{sd} \quad (1)$$

Where R_{sh} and R_{hd} are data rate for two hop transmission from source to relay and relay to destination respectively. R_{sd} is the direct supported data rate from source to destination. Every intermediate station which hears both RTS and CTS will calculate the value of R_{sh} and R_{hd} from its *CoopTable*.

D. Data Transmission Using Relay Station

In this proposed cooperative MAC, when a source has a frame for transmission it sends RTS frame using IEEE 802.11DCF. All other stations including the destination station shall receive this RTS frame. If the RTS-CTS are not activated by IEEE 802.11 DCF, this means that basic access mechanism has been enabled. This decision is made using RTS threshold. In basic access method, cooperative MAC mechanism is not triggered. After sending the RTS frame, the sender will wait for the CTS and HCTS (Helper Clear to Send) frames. Figure 6 describes time line of control frames and data frame transmission in our proposed cooperative MAC. If receiver station successfully receives the RTS it will reply by a CTS frame to the sender. Receiver indicates maximum supported direct data rate in the PLCP (Physical Layer Convergence Protocol) header of CTS frame. From CTS frame helper station comes to know direct supported maximum data rate between the source and destination.

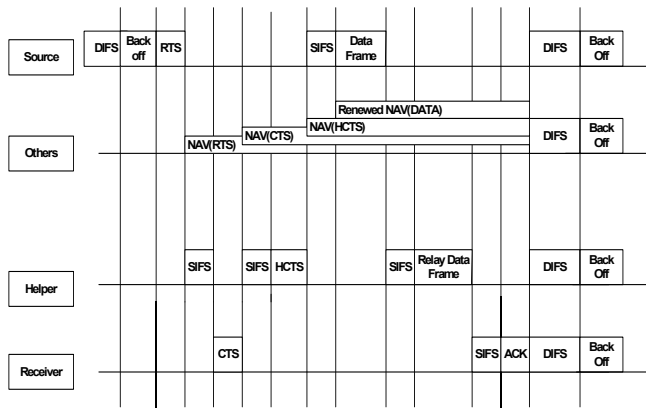


Fig. 6. RTS-CTS-HCTS control frame handshaking in the proposed Cooperative MAC

If the supported data rate falls below a threshold rate (2 Mbps for IEEE 802.11 b), it sends an HCTS frame after DIFS time interval. Upon receiving the HCTS frame, the sender sends data frame to the helper with transmission rate R_{sh} and then the relay station forward the data frame to destination with data rate R_{hd}.

TABLE 1. COOPMAC TABLE FORMAT

MAC Address (48 bits)	Data Rate(8 bits)	Time Stamp(8 bits)	RSSI(dBm)(8 bits)
Mac addresses of the adjacent Neighbors	Maximum Data rate	Time of the last RTS/CTS heard from the neighbor	RSS Value of the last Received RTS/CTS from the neighbor

If the sender does not receive any HCTS within DIFS interval time, the sender sends data frame directly to the destination. Most specifically a station declares itself as a helper if the following condition is satisfied.

$$\frac{8L}{R_{hs}} + \frac{8L}{R_{hd}} + 2T_{PLCP} + T_{HCTS} + 2T_{SIFS} + T_{DIFS} < \frac{8L}{R_{sd}} \quad (2)$$

Where, L is the length of the Data Frame. If this condition is satisfied, the value of duration field (DI) of data frame header gets minimized than that of without cooperation. Correspondingly, NAV value of waiting stations is set to new lesser value. Therefore, other stations get more frequent access to the channel. Duration field of control frames in proposed Cooperative MAC are shown in Table 2.

E. Collision Avoidance among Helper Stations

A helper station is selected in a distributed manner. If more than one station satisfies (2), each declares itself as the helper for that particular data transmission session. Hence, all of them try to send HCTS at the same time. Multiple HCTS induce collisions among different relay stations. Whenever an intermediate node receives both RTS and CTS from sender and receiver and it satisfies the condition (2), it senses the medium for DIFS time and it starts decrementing its back off counter. If a station finds the channel idle for DIFS time, it starts sending HCTS control frame to inform the source and receivers of its presence. All other intermediate stations that intended to be a helper finds the medium busy or their back off counter may not reach to zero. Therefore, they refrain from transmitting HCTS control frames. Sending station waits for DIFS time before transmitting a data frame.

TABLE 2. DURATION FIELD OF CONTROL FRAMES AND DATA FRAMES

Frame type	Duration Value
RTS	$3T_{SIFS} + T_{CTS} + T_{ACK} + T_{DATA}$
CTS	$2T_{SIFS} + T_{ACK} + T_{DATA}$
$HCTS$	$3T_{SIFS} + T_{ACK} + \frac{8L}{R_{hs}} + \frac{8L}{R_{hd}} + 2T_{PLCP}$
$Data\ Frame$	$2T_{SIFS} + T_{ACK} + \frac{8L}{R_{hs}} + \frac{8L}{R_{hd}} + 2T_{PLCP}$

V. ANALYTICAL MODELING

A. Throughput Analysis in Error Free Chanel

According to the discussion in section III and section IV, if we assume that are n number of stations equally distributed over a region and each station has an equal expected transmission probability with τ in randomly selected time slot. Let, P_{trans} denotes the probability that there is at least one transmission in randomly selected slot time

$$P_{trans} = 1 - (1-\tau)^n \quad (3)$$

Let, p denotes the probability that a collision occurs during a randomly selected slot time. Therefore, the value of τ and p can be found by solving the following nonlinear system given in [10]

$$\tau = \frac{2(i-2p)1-p^{m+1})}{(CW_{min}+1)(1-(2p)^{m+1})(1-p)+(1-2p)(1-p^{m+1})} \quad (4)$$

$$p = 1 - (1-\tau)^{n-1} \quad (5)$$

Where, CW_{min} is minimum contention window size, m denotes the maximum number of retry.

Let, P_s be the probability that a successful transmission happens that is the probability that a station transmit and other $(n-1)$ stations refrain from transmission in that given slot time. Therefore, P_s is a conditional probability conditioned on P_{trans}.

$$P_s = \frac{n\tau(1-\tau)^{n-1}}{P_{trans}} \quad (6)$$

Duration of a slot time= Fraction of idle time (no transmission) + fraction of time spent in collision+ fraction of time spent in a successful transmission.

Therefore, Throughput S can be derived using the following equation:

$$S = \frac{E[Payload\ in\ bits\ transmitted\ in\ a\ slot\ time]}{E[Duration\ of\ a\ slot\ time]}$$

$$S = \frac{P_b p_s E[L]}{(1-P_{trans})\pi + P_{trans}p_s T_s + P_{trans}(1-p_s)T_c} \quad (7)$$

Here, $\pi, T_c, and\ T_s$ is duration of an empty time slot, average time spent in collision in a time slot and medium is sensed busy due to a successful transmission of a frame in a time slot respectively. If RTS/ CTS access method is employed, T_s and T_c for IEEE 802.11 DCF MAC without cooperation and IEEE 802.11 DCF Cooperative MAC with our proposed helper selection method can be written as follows:

$$T_s^{Without\ CoOperation} = DIFS + T_{RTS} + T_H + 3SIFS + T_{ACK} + T_{CTS} + T_{E[P]}$$

$$T_C^{Without\ CoOperation} = DIFS + T_{RTS} \quad (8)$$

$$T_C^{With\ CoOperation} = DIFS + T_{RTS} + T_{HCTS}$$

$$T_s^{With\ CoOperation} = 2DIFS + T_{RTS} + 2T_H + 4SIFS + T_{ACK} + T_{CTS} + T_{HCTS} + \frac{T_{E[P]}}{R_{hs}} + \frac{T_{E[P]}}{R_{hd}}$$

Where, T_{RTS} , T_{ACK} , T_{CTS}, T_{HCTS}, $T_{E[P]}$ and T_H are the required transmission time for sending RTS, ACK, CTS, HCTS, Data Frame and Frame Header respectively from sender to receiver.

B. Throutput Analysis in Error Prone Environment

Throughput in (7) considers that frame error is caused due to collisions among the transmitted packets. If the transmission error is considered, the probability that a frame is corrupted due to collision or channel bit error can be written as

$$p = 1 - (1-\tau)^{n-1}(1-p_e)^{L_{payload}+L_{Header}} \quad (9)$$

Where p_e is the probability is denotes the channel bit error, $L_{payload}$ denotes length of the frame in bits, L_{Header} denotes the length of the frame header in bits. Now the equation (9) and equation (4) are nonlinear system. Numerical solution of (4) and (9) gives the value of p and τ.Frame Error rate can be written as

$$FER = 1 - (1-p_e)^{L_{payload}+L_{Header}} \quad (10)$$

Probability that a frame is successful given that there is at least one frame transmitted in a time slot with no frame error is

$$P_s = \frac{n\tau(1-\tau)^{n-1}}{P_{trans}}(1 - FER) \quad (11)$$

Probability that a transmitted frame received by the receiver with error can be written as

$$P_{ErrFrame} = P_s.FER \quad (12)$$

Therefore, Throughput in (7) can be re-written for error prone channel as follows

$$S = \frac{P_b p_s E[L]}{(1-P_{trans})\pi + P_{trans}p_s T_s + P_{trans}(1-p_s)T_c + P_{trans}.P_{ErrFrame}.T_{ErrFrame}} \quad (13)$$

Where $T_{ErrFrame}$ denotes the average time the medium is sensed busy due to corrupted frame transmission.

TABLE -3: SYSTEM PARAMETERS FOR MAC AND PHYSICAL LAYER

Packet Payload	8224 bits
MAC Header	224 bits
PHY Header	192 bits
ACK	112 bits +PHY Header
HCTS	112 bits +PHY Header
RTS	160 bits +PHY Header
CTS	112 bits + PHY Header
Channel Basic Data Rate	1 Mbps
Slot Time	20 µs
SIFS	10 µs
DIFS	50 µs

If RTS/CTS mechanism is employed, $T_{ErrorFrame}$ in error prone channel can be written as

$$T_{ErrorFrame}^{Without\ CoOperation} = DIFS + T_{RTS} + T_H + 3SIFS + T_{ACK} + T_{CTS} + T_{E[P]} \quad (14)$$

For the proposed cooperative MAC, $T_{ErrFrame}$ can be written as

$$T_{ErrFrame}^{With\ CoOperation} = 2DIFS + T_{RTS} + 2T_H + 4SIFS + T_{ACK} + T_{CTS} + T_{HCTS} + \frac{T_{E[P]}}{R_{hs}} + \frac{T_{E[P]}}{R_{hd}} \quad (15)$$

C. Result Discussion

In this paper IEEE 802.11b DSSS is used for the analysis of our proposed Cooperative MAC. In our analysis we have shown how our proposed CooPMAC and IEEE 802.11 DCF MAC throughput differs with the changes in frame size and number of stations. In figure 7, it shows that throughput of 1 Mbps data rate stations increases with the increase in frame size and number of stations in the wireless network.

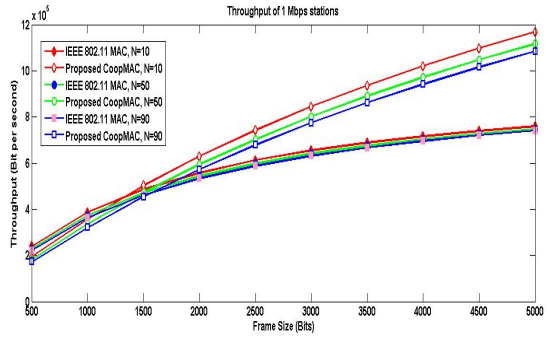


Fig. 7. Throughput versus Frame Size for 1 Mbps stations with different number of stations and frame sizes.

For number of stations, N=10, 50 and 90, our CooPMAC is beneficial if the frame sizes are larger than threshold 1300 bits, 1500 bits and 1600 bits respectively. It is due to the fact that extra overhead $DIFS + T_{HCTS} + SIFS$ incurred in CooPMAC for helper node selection. If the frame size increases, throughput efficiency of 1 Mbps stations increases dramatically in our CoopMAC in respect to IEEE 802.11 DCF MAC. In figure 8, it depicts that 2 Mbps stations achieve more throughput than IEEE 802.11 DCF MAC with higher frame size threshold. In both cases, slow stations (i.e. 1 Mbps and 2 Mbps in IEEE 802.11 b) are highly benefited if our proposed Cooperative MAC is applied.

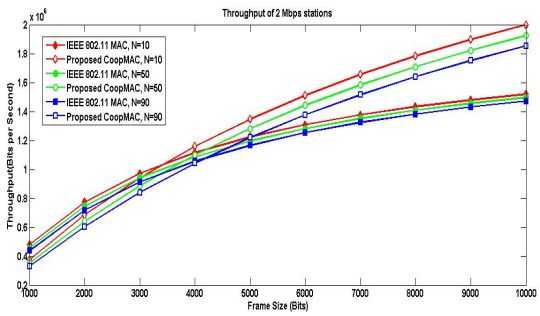


Fig. 8. Throughput Versus Frame size for 2 Mbps stations with different number stations

In figure 9, it is shown the variation of throughput achievement for different data rate stations in IEEE 802.11b with the increases wireless stations for a fixed frame size 8224 bits. It reveals that cooperative MAC is not beneficial for stations with data rates 5.5 Mbps and 11 Mbps. It indicates that extra overhead and delay induced in (9) is unnecessary for higher data rate stations (i.e. 5.5 Mbps, 11Mbps).Throughput of the 1 and 2 mbps stations with considering error prone channel are shown in fig. 10 and fig.11 respectively. As the received power at the destination station decreases with distance of signal has travelled.

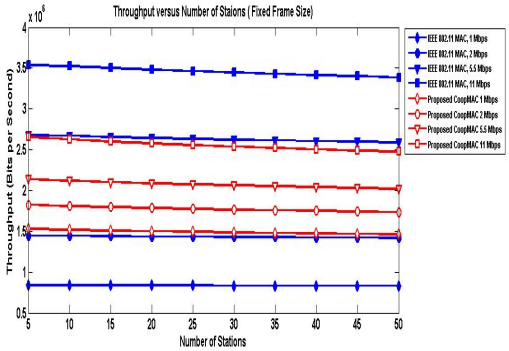


Fig. 9. Throughput of IEEE 802.11 b Stations

With the same transmit power; the relay station would receive the signal with more received power. BER would be lower for sender to relay and relay to receiver links than that of direct data transmission link.

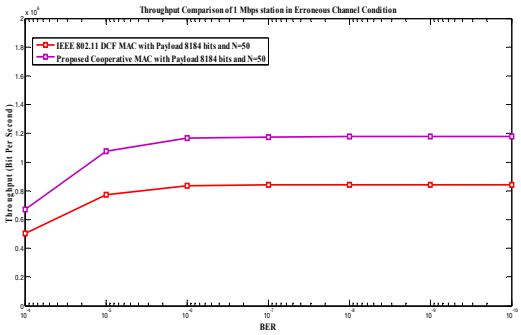


Fig. 10 Throughput comparison of 1 Mbps station with erroneous channel

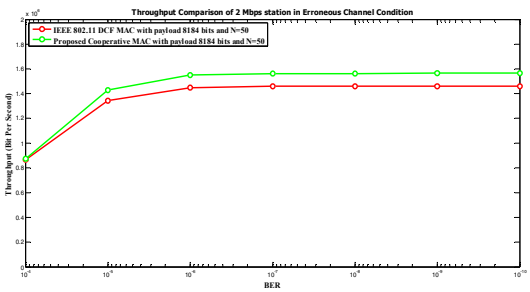


Fig.11 Throughput comparison of 2 Mbps station with erroneous channel

If the direct transmission (source–destination) has the BER at the destination 10^{-5} , we consider in our analysis that average BER for sender to relay and relay to destination links is little less than 10^{-5} with payload size of 8184 bits and 50 wireless stations.

VI. Conclusion and Future Works

We have studied IEEE 802.11 DCF MAC and our proposed cooperative MAC. Relay selection process is the corner stone in designing a cooperative MAC protocol. To facilitate the wireless stations to gain more access to wireless medium our protocol has minimized the overhead time in helper selection. The low rate stations speed up their transmissions through cooperation from helper node. Numerical analysis shows that lower rate stations gain more throughput efficiency than the higher rate stations. For higher rate stations initiating a cooperative transmission reduces their throughput than IEEE 802.11 DCF MAC. Higher rate stations need not initiate cooperative transmission.

As the helper station is selected based on the back off counter decrements, the best relay station among the best may not get selected. In future, we shall try to derive a strategy for selecting the best helper along with the analysis of the protocol using helper node mobility.

References

[1] J.N. Laneman, "Cooperative diversity: models, algorithms, and architectures", Cooperation in wireless networks: principles and applications, chapeter 1, Springer, 2006.

[2] A. Nosratinia, T. E. Hunter and A. Hedayat, "Cooperative communication in wireless networks" IEEE communication magazine, vol. 42, No.10, pp. 74-80, Oct. 2004

[3] H. Zhu and G. Cao, "rDCF: A relay enabled medium access control protocol for wireless ad hoc networks", IEEE transactions on mobile computing, vol. 5, pp. 1201-1213, 2006

[4] P. Liu, Z.Tao, and S.S.Panwar, "A cooperative mac protocol for wireless local area networks", IEEE journal on selected areas in communicatons, vol. 25, no.2, February, 2007.

[5] H. Shan, H.T. Cheng, W. Zhuang, "Cross layer cooperative MAC protocol in distributed wireless networks", IEEE transaction on wireless communications, vol. 10, no. 8, August, 2011

[6] X. He, F.Y. Li , "Cooperative medium access control in wireless networks : The two hop case" , In proceedings of IEEE wireless and mobile computing, networking and communications vol. 1 pp. 13-18, 2009

[7] S. Bharati, W.Zhuang " Performance analysis of cooperative adhoc MAC for vehicular networks", IEEE Globecom –wireless networking symposium , 2012

[8] S. Moh, C.Yu, "A cooperative diversity based robust MAC protocol in wireless ad hoc networks", IEEE transactions on parallel and distributed sytesms, vol. 22, no.3 March 2011.

[9] T. Zhou, H. Sharif, M. Hempel, P. Mahasukhon, W. Wang, T. Ma, "A novel adaptive distributed cooperative relaying mac protocol for vehicular networks",IEEE Journal on selected areas in communication, vol. 29, no.1 January 2011.

[10] P. Chatzimisios, V. Vitsas and A.C. Boucouvalas "Throughput and delay analysis of IEEE 802.11 protocol", In the proc. of the 5th IEEE international workshop on network appliances Liverpool John Moore's University, UK, pp. 168-174, Oct. 30-31, 2002

Bangla News Classification using Naive Bayes classifier

Abu Nowshed Chy
Dept. of Computer
Science & Engineering
University of Chittagong
Chittagong - 4331, Bangladesh
Email: nowshed@skeim.org

Md. Hanif Seddiqui
Dept. of Computer Science & Engineering
Bangabandhu Sheikh Mujibur Rahman
Science and Technology University, Gopalganj-8100
and University of Chittagong, Chittagong-4331
Email: hanif@cu.ac.bd

Sowmitra Das
Dept. of Computer
Science & Engineering
University of Chittagong
Chittagong-4331, Bangladesh
Email: sowmitra@skeim.org

***Abstract*—Web is gigantic and being constantly update. Bangla news in web are rapidly grown in the era of information age where each news site has its own different layout and categorization for grouping news. These heterogeneity of layout and categorization can not always satisfy individual user's need. Removing these heterogeneity and classifying the news articles according to user preference is a formidable task. In this paper, we propose an approach that provides a user to find out news articles which are related to a specific classification. We use our own developed web crawler to extract useful text from HTML pages of news article contents to construct a Full-Text-RSS. Each news article contents is tokenized with a modified light-weight Bangla Stemmer. In order to achieve better classification result, we remove the less significant words i.e. $stop-word$ from the document. We apply the naive Bayes classifier for classification of Bangla news article contents based on news code of IPTC. Our experimental result shows the effectiveness of our classification system.**

I. Introduction

The Web is rapidly moving towards a platform for mass collaboration in content production and consumption, and the increasing number of people are turning to online source for daily news. Bangla online newspaper started appearing at about the same time that the internet became public in Bangladesh. Every day, vast amount of Bangla news articles are created by the several news sites that exist in the World-Wide-Web and its rate increases exponentially. An online newspaper has many forms [1]. One form is electronic edition of the printed newspaper. The user can read the online edition similar to a paper edition; there is no categorization, neither with respect to content nor with respect to layout. Another form of online newspaper is news website, which enables the user browsing in menus that are organized in subject categories and sub-categories. Common to most of the above forms of online newspapers is that the user is assumed to read the news from a computer screen, while connected via the Internet to a certain news provider. However, these services might not be sufficient for many readers and reading situations.

A lot of newspaper reader like to view and analyze news from various news sources. Many times, readers are interested only in the news articles of their categories of interest [2]. Therefore,the users have to scan through all the news articles of several news sites in order to get the articles of his/her interest. For example, a user interested in sports related news has to go through all the news articles from various news sites and their time spent in analyzing news from the multiple sources which is tedious. So a reader would prefer a system that would gather news articles from various sources; which is accessible to all the time and anywhere, also from a mobile reading device. A reader would like to visit a unique newspaper that includes the articles from various favorite sources, arranged and presented in an order that best fits her interests and reading habits.

Moreover, International Press Telecommunication Council (IPTC) publishes as standard of news classification, called as news code taxonomies. The IPTC creates and maintains sets of concepts, called a controlled vocabulary or a taxonomy to be assigned as metadata values to news objects like text, photographs, graphics, audio and video files and streams. This allows for a consistent coding of news metadata across news providers and over the course of time [1].

We hardly find the IPTC standard classification at most of the online Bangla news media. Therefore, our attempt is to classify the online news article contents to satisfy users query.

Web pages have their own underlying embedded structure in the HTML language [3]. They typically contain noisy content such as advertisement banner and navigation bar. If a pure-text classification method is directly applied to these pages, it will incur much bias for the classification algorithm, making it possible to lose focus on the main topics and important content. Thus, a critical issue is to design an intelligent preprocessing technique to extract individual news document from a Web page. Our web crawler crawl each news site and extract the individual news document of specific date from that news site as a Full-Text-RSS file. The indidividual document is then tokenized with a light-weight bangla stemmer and the less significant words, which are called stop-words, are removed. Finally, classification is performed to each news item based on IPTC news code taxonomy. The basic news classification problem can be formulated as follows:

Given a training set of documents $D_{train} = (d_1, l_1)....(d_n, l_n)$ of labeled news documents where each document d_i belongs to a document set D and the label $l_i = l_i(d_i)$ of d_i is within a manually predefined set of categories $C = c_1, ..., c_m$ of IPTC news code taxonomy. The goal in news document categorization is to devise a learning algorithm that given the training set D_{train} as input will generate a classifier $h : D- > C$ that will be able to

[1] http://www.iptc.org/site/NewsCodes/

accurately classify unseen documents from D [4].

The rest of this paper is structured as follows: **Section II** describes the related work while the foundamental concepts of news code of IPTC is articulated in **Section III**. We introduce our approach in **Section IV**. **Section V** includes experiments and evaluation to show the effectiveness of our proposed approach. Concluded remarks and some future directions of our work is described in **Section VI**.

II. Related Work

The task of news classification is to automatically classify news documents into predefined classes based on their content. A number of statistical and machine learning techniques has been developed for news classification in different language. In [5], KNN classification algorithm is used to predict the test samples category according to the K training samples which are the nearest neighbors to the test sample and judge it to that category which has the largest category probability. [6] propose Weight Adjusted k-Nearest Neighbor (WAKNN) classification algorithm that is based on the k-NN classification paradigm. In WAKNN, the importance of each word in the classification of a training document set is learned and the weight vector reflecting this importance is maintained. The weight vector is used in the similarity measure computation such that important words contribute more in the similarity measure.

Naive Bayes (NaiveBayes) probabilistic classifier is often used in text classification applications and experiments because of its simplicity and effectiveness which basic idea is to use the joint probabilities of words and categories to estimate the probabilities of categories given a document [7]. The naive part of such a model is the assumption of word independence. However, its performance is often degraded because it does not model text well, and by inappropriate feature selection and the lack of reliable confidence scores. In [8], Karl-Michael Schneider address these problems and show that they can be solved by some simple corrections. The paper [9] shows that the accuracy of a NaiveBayes classifier can be significantly improved by taking advantage of a hierarchy of classes. They adopt an established statistical technique called shrinkage that smoothes parameter estimates of a data-sparse child with its parent in order to obtain more robust parameter estimates and deleted interpolation that smoothes n-grams in language modeling.

In text classification, Support Vector Machine (SVM) classifier transforms the original data in a higher dimension, from where it can find a hyperplane for separation of the data using essential training instances called support vectors and thus classify the document [10]. But SVM classifier provide excellent precision with poor recall. In paper [11] James G. Shanahan and Norbert Roma describe an automatic process for adjusting the thresholds of generic SVM to improve recall. [12] used n-gram (n character slice of a longer string)based classifier algorithm for Bangla text categorization.

An n-gram is a subsequence of n-items in any given sequence. In computational linguistics n-gram models are used most commonly in predicting words (in word level n-gram) or predicting characters (in character level n-gram). Typically, n is fixed for a particular corpus of documents and the queries made against that corpus where corpus is a huge text. In n-gram based text classification, Johannes Furnkranz [13] shows that after the removal of stop words, word sequences of length 2 or 3 are most useful.

Using longer sequences reduces classification performance. [14] used Linear Least Squares Fit (LLSF) classifier that use a large collection of human-assigned matches between texts and categories as a ”training set”, to compute word-to-category connections that enables us to weight these connections in such a way as to optimally fit the training set and to probabilistically capture the correct categories for arbitrary texts. [2] used Vector-Space-Model classifier which basic idea is to construct a prototype vector per category using a training set of documents. Given a category, the vectors of documents belonging to this category are given a positive weight, and the vectors of remaining documents are given a negative weight. By summing up these positively and negatively weighted vectors, the prototype vector of this category is obtained. Rodriguez et. al. described an approach that integrates WordNet with two training approaches (Rocchio - relevance feedback and Widrow Hoff - machine learning) through the Vector Space Model which improve their performance [15].

Decision tree (DTree) algorithms select informative words based on an information gain criterion, and predict categories of each document according to the occurrence of word combinations in the document [16]. [17] described a fast decision tree induction algorithm especially suited to text data and a rule simplification method that converts the decision tree into a logically equivalent rule set.

Corner classification (CC) network is a kind of feed forward neural network for instantly document classification. To classify text object instantly, new training algorithm-TextCC, for feed forward neural network is presented in [18]. since textural data has a very high-dimension feature space text classification model using an artificial neural network as the text classifier scalability is poor if the neural network is trained using the raw feature space.

Another paper propsed four dimensionality reduction techniques and show that Principal Component Analysis (PCA) was found to be the most effective in reducing the dimensionality of the feature space [19].

Some researchers comparing the performance of two or more indidividual classifier on the same data set to show that which classifier perform well on what kind of data set [4] [20]. In the context of combining multiple classifiers for text clssification, a number of researchers have shown that combining multiple classifiers can improve classification accuracy [21] [22]. An Evaluation of number of Statistical Approaches to Text Categorization is given in [23].

III. News Code of IPTC

To promote the ease of interchange of news items, The International Press Telecommunication Council (IPTC), an international organization that is primarily focused on developing and publishing Industry Standards for the interchange of news data, has provided numerous categorization schemes aimed to standardize coding of various aspects of news related metadata. The whole product is called NewsCodes. A NewsCode is a

single code representing a concept which is used to categorize news content. Many of these codes can make a set for a specific use, such a set is branded as NewsCodes by the IPTC.

Typically NewsCodes are language agnostic and as each code has an explicit and comprehensive definition so one can easily share not only the codes but also their semantics. More generic terms used are vocavulary, taxonomy, topicset and so on. Vocabulary is a set of codes. Can be either controlled or uncontrolled. Controlled Vocabulary managed by some authority (e.g. a person or an organisation), employing some mechanism (e.g. an XML Schema, IPTC G2 KnowledgeItem) to maintain this set. Each code in a controlled vocabulary represents a concept. Concept means anything that one may wish to refer to and may be represented by one or more codes.

Collection of concept with associated code is called taxonomy.A taxonomy may support typed relationships between concepts. Such a taxonomy is sometimes known as an ontology or thesaurus. The IPTC groups the taxonomies into four main areas: **Descriptive NewsCodes** - is a group of taxonomies to describe the content of news items properly. **Administrative NewsCodes** - is a group of taxonomies for proper administration of news items. **Transmission NewsCodes** is a group of taxonomies with controlled values for the transmission process. **Exchange Format NewsCodes** - is a group of taxonomies with values to support specific functionalities of the different IPTC news exchange format standards.

IV. Our Approach

The goal of news classification is to assign categories to a news document according to the content of the document. In this regard, we perform a number of steps as follows:

1. A number of online news sites are selected based on type and popularity.
2. By using our own RSS crawler we crawl each sites and generate Full-Text-RSS that contains title, link, description and date.
3. Then we use a RSS parser to extract individual news document from the RSS feed.
4. After getting idividual document we process the documents text by several techniques such as tokenization, punctuation removal, digit removal and so on.
5. We build a light bengali lexicon i.e. begali dictionary that help us in stemming.
6. We establish a light bengali stemmer to identify the stem of each word.
7. We then perform some advanced text processing task i.e, single letter words removal, stop word removal.
8. Finally, we consider each document as a word-vector and supervised learning is conducted to assign multiple relevant classes.
9. Results of the supervised learning task are applied to new documents and we categorize the documents based on Naive-Bayes classifier.

The overview of our approach is depicted in Fig. 1.

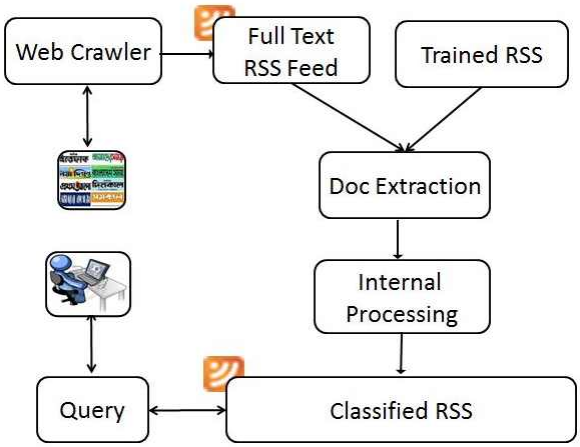


Fig. 1. Our News Classification System Overview

A. Construction of Full-Text-RSS

To create Full-Text-RSS, we extract the Web news article contents from news sites. Usually, the extraction process includes two steps. Firstly, the news sites are crawled on Breadth-First-Search method to collect the news pages. Secondly, the news article contents are extracted from news pages and generate RSS file. As the news sites comprise different kinds of Web pages, there are many non-news pages such as the advertisement, related stories and comments. In order to recognize and extract the parts of news article contents from the full text of news pages, wrappers are generated based on the analysis of layout of news pages by many extraction methods. Web page layout is the style of graphic design in which text or pictures are set out on a Web page. The different news sites use the different news page layout, and each news site uses more than one layout. Inspite of the heterogeneous layouts we analyze the news pages to detect an actual news url and the title, news-body and publication date is extracted. We observe that a page containing smaller number of links is likely to be a detail of news article content. Once we find an actual news url, extracting title, news-body and the publication date is relatively easier to extract.

B. News Data Preprocessing

An RSS parser extracts individual news document from the Full-Text-RSS feed. After the extraction of news document the preprocessing step is initiated with tokenization.

1) Tokenization: Tokenization is the process that aims to fracture the stream of characters into tokens delimited by white space, tab, new line and so on. As we are working on Bengali news document, there are lots of Bengali punctuations in the document. Moreover news document may contain Bengali as well as English digits. As meaningful Bengali words do not contain these characters we remove these. We also eliminte the single letter word from the document in this stage. Then the documents are nothing but a bag-of-words. As Bengali is a highly inflected language with relatively free or pragmatically free word order i.e., Bengali (verb, noun, adjective) words are inflected forms of roots we perform stemming on the tokenized words.

```
Algo. bangla_stemmer (word)

for each word ∈ document
   dictionary_checkers(word)
   if word ∈ dictionary
     stem=word
   else stem1=stemming(word)
      if stem1 ∈ dictionary
        stem=stem1
      else stem=stemming(stem1)
```

Fig. 2. Pseudo code of how our bengali stemmer works.

2) Stemming: Stemming is an operation that splits a word into the constituent root part and affix without doing complete morphological analysis. Terms with common stems tend to have similar meaning, which makes stemming an attractive option to increase the performance of spelling checkers and other information retrieval applications where morphological analysis would be too computationally expensive. Another advantage of stemming is that it can drastically reduce the dictionary sized used in various NLP applications, especially for highly inflected languages [24].

The simplicity of languages like English, and other English-Like Languages has enabled the development of stemmers for them. Moreover, these languages being web-resource enriched, make the work of the researchers easy. However, the languages which are highly inflectional and resource poor at the same time have always been treated with negligence, especially Bangla. Though Bangla is the fourth most widely spoken language in the world. In the written form of Bangla there are 11 vowels and 39 consonants. Moreover, there are 10 short forms of vowels called vowel modifiers (i.e. Kar), 7 short forms of consonants called consonant modifiers (i.e.Fala). The last two decades has witnessed an immense escalation of Bangla web and digital text contents and is having an exponential growth rate. This has enhanced the need for the development of highly efficient Information Retrieval (IR) systems and consequently good stemmers [25].

By considering the above facts we design a Bangla stemmer to improve the performance of our system. Moreover we create a light Bangla dictionary that help us stemming a word. To achieve this we write a parsing program that extract Bangla word from a online bilingual dictionary. There are much inflected word on this dictionary so we remove these inflection from the dictionary. This dictionary contains about 15000 words that frequently used in online Bangla news. The dictionary plays an important role while stemming a word. The algorithm in Fig. 2 demonstrate how our Bangla stemmer works with the help of light Bangla dictionary.

3) Stop Word Removal: Statistical analysis through documents shows that some words have quite low frequency, while some others act just the opposite. The characteristic of these common words is that they carry almost no significant information to the document. Instead, they are used just because of grammar. We usually refer to this set of words as stop words [26].

For the purpose of this research we define a general stop word list. These non-significant words represent noise, and may actually damage the classification performance because they do not discriminate between relevant and non-relevant documents. Secondly, we expect to reduce the size of the considerable vector elements.

In establishing a general stop word list for Bangla, we follow several steps. At first, with the help of our RSS crawler we collect 3142 distinct news document as RSS file from a popular online Bangla newspaper prothom-alo.com. Each document is then tokenized and performed stemming. Finally, a simple feature selection method, "inverse document frequency (IDF)", is applied to determine stop-words, the less significant words in identifying documents. We sorted all the word according to their IDF value in ascending order and we consider the large stop word list (300 most frequently words) as a trend in information retrieval system over time. We inspect our system by considering other values of stop word list. We consider the first 300 words as stop words. Moreover, we included some other widely used non-information-bearing words in the stop word list manually.

C. Classification Algorithms

In Bangla news classification, we use the Naive Bayes Classifier.

1) Naive Bayes classifier: Naive Bayes (NB) probabilistic classifier algorithm was first proposed and used for text categorization task by D.Lewis [7]. NB is based on the bayes theorem in the probabilistic frame work. The basic idea is to use the joint probabilities of words and categories to estimate the probabilities of categories given a document. The naive part of such a model is the assumption of word independence. The simplicity of the assumption of word independence makes the computation of the Naive Bayes classifier far more efficient than the exponential complexity of non-naive Bayes approaches because it does not use word combinations as predictors.

When the NB classifier is applied in Text Classification problem we use the following equation:

$$p(class \mid doc.) = \frac{p(class).p(doc. \mid class)}{p(doc.)} \tag{1}$$

where $p(class \mid doc.)$ is the probability that a given document D belongs to a given class C. $P(doc.)$ is the probability of a document, we can notice that $p(doc.)$ is a Constance divider to every calculation, so we can ignore it. $P(class)$ is the probability of a class (or category), we can compute it from the number of documents in the category divided by documents number in all categories. $p(doc. \mid class)$ represents the probability of document given class and documents can be modelled as sets of words, thus the $p(doc. \mid class)$ can be written like:

$$p(doc. \mid class) = \prod_i p(word_i \mid class) \tag{2}$$

So, we can rewrite the equation(2) as:

$$p(class|doc.) = p(class). \prod_i p(word_i|class) \tag{3}$$

Where $P(word_i \mid class)$ is the probability that the i-th word of a given document occurs in a document from class, When the size of the training set is small, the relative frequency estimates of probabilities, $p(word_i \mid class)$, will not be reasonable; if a word never appears in the given training data, its relative frequency estimate will be zero. Instead, we applied the Laplace law of succession to estimate $p(word_i \mid class)$. The estimate of the probability $p(word_i \mid class)$ is therefore written as:

$$p(word_i \mid class) = \frac{T_c t + \lambda}{N_c + \lambda V} \tag{4}$$

Where $T_c t$ is the number of times the word occurs in that category C. N_c is the number of words in category C. V is the vocabulary size. λ is the positive constant, usually 1, or 0.5 to avoid zero probability [27] [28].

V. Experiments and Evaluation

in this section, we present the evaluation of our aproach and explain the implementation of our system

A. Evaluation Metrics

To evaluate the systems classification accuracy, i.e. the proportion of correctly classified news document, we used common Information Retrieval (IR) performance measures, i.e. precision and recall to evaluate the system.

Suppose that the number of the documents which are in C_j category in fact and also the classifier judge them to C_j category is a; the number of the documents which are not in C_j category in fact, however the classifier judge them to C_j category is b; the number of the documents which are in C_j category in fact, however the classifier do not judge them to C_j category is c; the number of the documents which are not in C_j category in fact and also the classifier do not judge them to C_j category is d [5]. That is, once we get contingency table, we can define precision and recall as follows:

Precision, P: Precision is the ratio of the number of documents which judge correctly by classifiers to the number of documents which classifiers judged to this category, so the precision of C_j is defined as following:

$$Precision, P = \frac{a}{a+b} \tag{5}$$

Recall, R: Recall rate is the ratio of the number of documents which judge correctly by classifiers to the number of documents which are this category in fact, so the recall rate of C_j is defined as following:

$$Recall, R = \frac{a}{a+c} \tag{6}$$

B. Quantitative Analysis

In the taxonomy of IPTC, there are several hundreds of concepts for news subject codes. However, we consider only the upper level 34 concepts such as 'religion and belief', 'social problem', 'agriculture', 'crime, law and justice', 'sport event', 'politics', 'election', 'government', 'transport accident', 'environmental pollution', 'natural disaster', 'health' and so on. We have more than seven thousand news topics extracted from the Daily Prothom Alo. However, six hundred twenty eight (628) news items are manually tagged with multiple classification codes.

Our approach starts learning with 208 news items to classify the other 416 news items. We evaluate our approach with the TREC [29] evaluation technique to produce Recall-Precision graph depicted in Fig. 3.

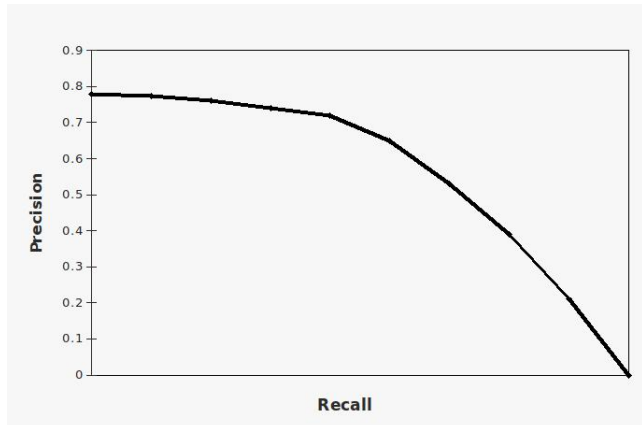


Fig. 3. Recall-Precision Graph of Our Bangla News Classification Approach

We are still on the way of evaluating news document classification with more lower level classification concepts those have some sort of overlapping characteristics.

VI. Conclusion and Future Direction

The growing domain of online newspaper presents a rich area, which can benefit immensely from automatic classification approach. In this paper we present a system of automatically classifying Bangla News documents. This system provides users with efficient and reliable access to classified news from different sources. It achieves a high accuracy of classifications with the possibility that one story be classified into more than one category. We used the Naive Bayes algorithm which is based on probabilistic framework to handle our classification problem. Bangla stemmer and stop word removal play an important role to achieve higher accuracy in Bangla news document classification.

Our future target is to consider the full set of IPTC news subject code for classification and to compare with other techniques such as KNN,SVM on Bangla news document classification) on our system.

References

[1] L. Tenenboim, B. Shapira, and P. Shoval, "Ontology-based classification of news in an electronic newspaper," 2008.

[2] B. Pendharkar, P. Ambekar, P. Godbole, S. Joshi, and S. Abhyankar, "Topic categorization of rss news feeds," *Group*, vol. 4, p. 1, 2007.

[3] D. Shen, Z. Chen, Q. Yang, H. Zeng, B. Zhang, Y. Lu, and W. Ma, "Web-page classification through summarization," in *Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval*. ACM, 2004, pp. 242–249.

[4] A. Kumar and M. Jain, "Text classification of news articles."

[5] Y. Zhou, Y. Li, and S. Xia, "An improved knn text classification algorithm based on clustering," *Journal of Computers*, vol. 4, no. 3, pp. 230–237, 2009.

[6] E. Han, G. Karypis, and V. Kumar, "Text categorization using weight adjusted k-nearest neighbor classification," *Advances in Knowledge Discovery and Data Mining*, pp. 53–65, 2001.

[7] D. Lewis, "Naive (bayes) at forty: The independence assumption in information retrieval," *Machine Learning: ECML-98*, pp. 4–15, 1998.

[8] K. Schneider, "Techniques for improving the performance of naive bayes for text classification," *Computational Linguistics and Intelligent Text Processing*, pp. 682–693, 2005.

[9] A. McCallum, R. Rosenfeld, T. Mitchell, and A. Ng, "Improving text classification by shrinkage in a hierarchy of classes," in *Proceedings of the Fifteenth International Conference on Machine Learning*, 1998, pp. 359–367.

[10] C. Ee and P. Lim, "Automated online news classification with personalization."

[11] J. Shanahan and N. Roma, "Improving svm text classification performance through threshold adjustment," *Machine Learning: ECML 2003*, pp. 361–372, 2003.

[12] M. Mansur, N. UzZaman, and M. Khan, "Analysis of n-gram based text categorization for bangla in a newspaper corpus," *Center for Research on Bangla Language Processing, BRAC University, Dhaka, Bangladesh*, 2005.

[13] J. Fürnkranz, "A study using n-gram features for text categorization," *Austrian Research Institute for Artifical Intelligence*, 1998.

[14] Y. Yang and C. Chute, "An example-based mapping method for text categorization and retrieval," *ACM Transactions on Information Systems (TOIS)*, vol. 12, no. 3, pp. 252–277, 1994.

[15] M. Rodriguez, J. Hidalgo, and B. Agudo, "Using wordnet to complement training information in text categorization," in *Proceedings of 2nd International Conference on Recent Advances in Natural Language Processing II: Selected Papers from RANLP*, vol. 97, 2000, pp. 353–364.

[16] D. Lewis and M. Ringuette, "A comparison of two learning algorithms for text categorization," in *Third annual symposium on document analysis and information retrieval*, vol. 33, 1994, pp. 81–93.

[17] D. Johnson, F. Oles, T. Zhang, and T. Goetz, "A decision-tree-based symbolic rule induction system for text categorization," *IBM Systems Journal*, vol. 41, no. 3, pp. 428–437, 2002.

[18] Z. Zhang, S. Zhang, E. Chen, X. Wang, and H. Cheng, "Textcc: new feed forward neural network for classifying documents instantly," *Advances in Neural Networks–ISNN 2005*, pp. 811–811, 2005.

[19] S. Lam and D. Lee, "Feature reduction for neural network based text categorization," in *Database Systems for Advanced Applications, 1999. Proceedings., 6th International Conference on*. IEEE, 1999, pp. 195–202.

[20] A. Babu and P. Kumar, "Comparing neural network approach with n-gram approach for text categorization," *International Journal on Computer Science and Engineering*, vol. 2, no. 1, pp. 80–83, 2010.

[21] Y. Bao and N. Ishii, "Combining multiple k-nearest neighbor classifiers for text classification by reducts," in *Discovery Science*. Springer, 2002, pp. 313–348.

[22] S. Cho and J. Lee, "Learning neural network ensemble for practical text classification," *Intelligent Data Engineering and Automated Learning*, pp. 1032–1036, 2003.

[23] Y. Yang, "An evaluation of statistical approaches to text categorization," *Information retrieval*, vol. 1, no. 1, pp. 69–90, 1999.

[24] M. Islam, M. Uddin, M. Khan *et al.*, "A light weight stemmer for bengali and its use in spelling checker," 2007.

[25] B. Das, "Development of bengali language stemmer," Ph.D. dissertation, Indian Institute of Technology.

[26] F. Zou, F. Wang, X. Deng, and S. Han, "Automatic identification of chinese stop words," *Research on Computing Science*, vol. 18, pp. 151–162, 2006.

[27] F. Thabtah, M. Eljinini, M. Zamzeer, and W. Hadi, "Naïve bayesian based on chi square to categorize arabic data," in *proceedings of The 11th International Business Information Management Association Conference (IBIMA) Conference on Innovation and Knowledge Management in Twin Track Economies, Cairo, Egypt*, 2009, pp. 4–6.

[28] E. Frank and R. Bouckaert, "Naive bayes for text classification with unbalanced classes," *Knowledge Discovery in Databases: PKDD 2006*, pp. 503–510, 2006.

[29] E. Voorhees and D. Harman, *TREC: Experiment and Evaluation in Information Retrieval*. MIT Press Cambridge, 2005, vol. 63.

A Smart User Interface for Efficient Resource Allocation for Arsenic Mitigation in Bangladesh

Mohammad Mamun Elahi, Muhammad Ilias Amin, Mohammad Nazrul Islam, Mohammad Masudul Haque
Department of Computer Science and Engineering
United International University
Dhaka, Bangladesh.
mmelahi@cse.uiu.ac.bd, ilias.amin@gmail.com, nazrul.ois@gmail.com, mail_masud@yahoo.com

***Abstract*— Arsenic contamination of groundwater in many nations including Bangladesh shows that this is a global problem. Because of the delayed health effects, poor reporting, and low levels of awareness in some communities, the extent of the adverse health problems caused by arsenic in drinking-water is at alarming level in Bangladesh. Also, allocating resources such as tube wells efficiently and effectively to mitigate arsenic hazard is a challenging task in Bangladesh. To allocate resources based on different arsenic hazard parameters, we have developed a Decision Support System that enables the user to observe the effect of allocation policy both in tabular and spatial format using statistical models. We have also developed an algorithm for optimal allocation of resources. A Smart User Interface is designed for the users so that they will find an interactive, user-friendly, intelligible, logical, clear, and sound environment to work with. Finally, we have analyzed and demonstrated the efficacy of our algorithm graphically.**

Keywords*— *Arsenic hazard, Decision Support System, Geographic Information System, and Vulnerability Index.

I. INTRODUCTION

About 97% of Bangladeshi people have access to a water point and unfortunately only 76% of these can now be considered safe considering permitted level of arsenic in drinking water. The United Nations World Health Organization calls Bangladesh's arsenic crisis "The largest mass poisoning of a population in history." Drinking-water progress towards the Millennium Development Goal (MDG) target is in progress but yet insufficient for Bangladesh [1]. That is why; arsenic mitigation both in short term and long term basis have become is a burning issue for Bangladesh.

The Arsenic Decision Support System (ADSS) is a Geographic Information System (GIS) tool developed by The CEGIS (Center for Environmental and Geographic Information Services) as a pilot project for analyzing arsenic situation of an area using statistical model [6] [11]. The study was conducted in three upazilas namely Muradnagar of Comilla district, Sirajdikhan of Munshiganj district and Bhanga of Faridpur district. CEGIS has collected data in 2001 and 2005. In 2005, all tube wells were not tested, so there were some changes, and further adjusted according to the changes. The CEGIS ADSS framework computes a single attribute named Vulnerability Index (V.I.) by some arithmetical formula. ADSS works as a decision support system for analyzing existing arsenic data and viewing output both in tabular and spatial format. To extend the existing ADSS of CEGIS, we take the opportunity to extend their project by making "An Integrated Smart User Interface" including decision making capabilities into it.

Our proposed Smart User Interface is designed for the users, where they will find an interactive, user-friendly, intelligible, logical, clear, and sound environment to work with. Here, in this system we have used all the field level data. The system is capable of comparing and differentiating between (Old VI and New VI) individually and also group-wise by just using a single click that reduces the users valuable time, cost and energy. The system will work just like a "One Stop Solution", where users will find all of their related queries instantly without any hassle.

The paper is organized as follows. Section II provides present arsenic situation in Bangladesh. Section III describes a decision support system on top of which we have implemented our system. Section IV describes the proposed resource distribution algorithm in details. Section V illustrates the smart interface and shows how both tabular and spatial outputs can be obtained by changing parameters. Finally, section VI ends the paper with concluding remarks.

II. PRESENT ARSENIC SITUATION IN BANGLADESH

The use of groundwater in Bangladesh for potable water supply and irrigation has rapidly increased since 1970. Arsenic contamination was first detected in 1993 by the Department of Public Health Engineering (DPHE), and its presence was subsequently confirmed after 1995 in numerous shallow and deep wells in different parts of the country.

According to the British Geological Survey (BGS), "the groundwater Arsenic (As) problem in Bangladesh arises because of an unfortunate combination of three factors: a source of As (As is present in the aquifer sediments), mobilization (As is released from the sediments to the groundwater) and transport (As is flushed in the natural groundwater circulation)" [9].

Bangladesh is one of the most heavily populated countries in the world and its drinking water is collected from different

sources, such as surface water, rain water and ground water. Surface water of Bangladesh is extremely contaminated with fecal bacteria, residues from fertilizers and pesticides. About 97% of the population of Bangladesh relies on ground water for drinking purpose according to "National policy for Arsenic mitigation 2004" report [2]. In Bangladesh, according to the Government's water quality standard, the permissible level of arsenic in drinking water is 50µg/L [3]. Ground water used for drinking in many areas of Bangladesh has been reported as contaminated by arsenic above the permissible level [1]. According to a survey conducted by BBS, Ministry of Planning, Government of Bangladesh and UNICEF in 2009, the population exposed to arsenic is estimated to 17.9 million [4]. Thousands of people have already been diagnosed with poisoning symptoms, even though much of the at-risk population has not yet been assessed for arsenic-related health problems, some of them even are not aware of arsenic contamination.

III. Arsenic Decision Support System

A. Decision Support System

A Decision Support System (DSS) is a computer-based information system that supports business or organizational decision-making activities. It serves the management, operations, and planning levels of an organization and help to make good decisions, which may be rapidly changing and not easily specified in advance. It includes knowledge-based systems. A properly designed DSS is an interactive software-based system intended to help decision makers compile useful information from a combination of raw data, documents, personal knowledge, or business models to identify and solve problems and make decisions. DSS is built in a variety of ways to classify data.

B. Components of a Decision Support System

Three fundamental components of DSS architecture are: Database, Model and User Interface.

- Database: A database is the information related to a particular organization that is processed and stored by a Database Management System (DBMS). DBMS is a software package with computer programs that control the creation, maintenance such as browsing, updating, deleting and inserting of a database.
- Model: While the quality and reliability of modeling tools and the internal architectures of DSS is important, the most crucial aspect of DSS is, by far, the user interfaces. The most important result of a session with a DSS is insight into the decision problem. In addition, when the system is based on normative principles, it can play a tutoring role; one might hope that users will learn the domain model and how to reason with it over time, and improve their own thinking [8]. The decision context and user criteria are specified using Model.
- User Interface: User Interface abbreviated as UI, the junction between a user and a computer program. An interface is a set of commands or menus through which a user communicates with a program. The user interface is one of the most important parts of any program because it determines how easily you can make the program do what you want. A good user interface to DSS should support model construction and model analysis, reasoning about the problem structure in addition to numerical calculations and both choice and optimization of decision variables.

C. Overview of the Arsenic Decision Support System

The Arsenic Decision Support System (ADSS) has been developed for visualizing arsenic situation, finding the suitable location to replace unusable tube wells (those with level of arsenic more than permissible limit, hence, marked as red tube well) in order to achieve the best result for reducing arsenic hazard of an area. To start with ADSS, let us first discuss about ADSS input data, which includes (1) Arsenic data (Arsenic Concentration, Number of people using tube wells), (2) Safe water option data (Pond sand filter, dug well, rainwater harvesting), (3) Arsenic related socio-economic data (drinking water sources, population awareness, and arsenic patients). ADSS inputs and calculates all data based on some administrative boundaries in Bangladesh. The lowest boundary is a Mouza under a particular union, which is under a particular thana. All parameters are calculated Mouza-wise.

The following figure (Fig. 1) depicts the process flow of the ADSS:

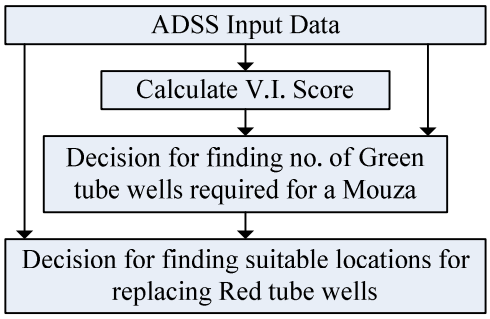


Fig. 1. Process flow of ADSS system

When all the input parameters are available, the framework can produce one of the major ADSS primary outputs, Vulnerability Index (VI), which is calculated from some pre-defined formulas based on domain knowledge, mathematical knowledge and statistical knowledge. This index represents the arsenic situation of a Mouza and can be compared with other Mouzas. Using VI information, ADSS helps to find number of new tube wells required for a Mouza and finally, tells the user where to place those tube wells.

Since ADSS resource distribution is based on their vulnerability, which means more vulnerable Mouzas are provided more tube wells then the less vulnerable Mouzas. Fig. 2 depicts the distribution of 105 tube wells among 7 Mouzas. X-axis shows the Mouza VI, and Y-axis shows number of tube wells to be distributed. It is to be noted that the Conventional distribution does not count or prioritize vulnerable Mouzas whereas; ADSS resource distribution method has better distribution according to their VI score.

Vulnerability Index: The Vulnerability Index (VI) of a Mouza is a numerical value, which represents the arsenic

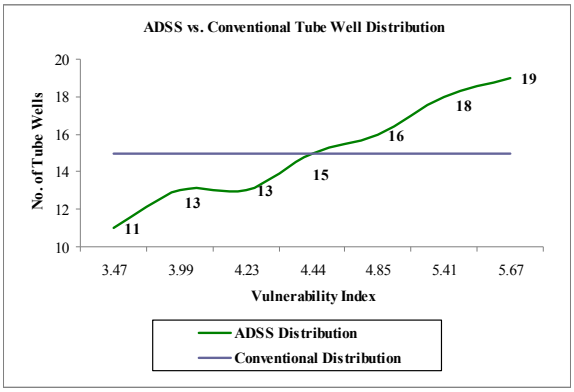


Fig. 2. ADSS vs. Conventional Tube Well Distribution

situation of that Mouza under a thana or a union. ADSS can calculate VI score in two different ways: absolute VI score represents vulnerability in a Mouza irrespective of other Mouzas in a thana or union. On the other hand, relative VI score enables us to compare among the Mouzas in a thana or union. However, the greater the VI value, the greater the arsenic risk. VI is calculated mainly from four main components: arsenic hazard, resources, coping capacity, and health hazard. The main components consist of other sub parameters. All these parameters (10 parameters) are listed in following table (Table I).

TABLE I. VI COMPONENTS (ADSS INPUT)

Parameter Name	Parameter Weight	Sub-Parameter Name	Sub-Parameter Weight
Arsenic Hazard Component	W_P	Percentage of red tube well (A)	W_A
		Average arsenic concentration (B)	W_B
		Percentage of population using red tube well (C)	W_C
Resource Component	W_Q	Percentage of green tube well (D)	W_D
		Average distance from usable pond (E)	W_E
		Average distance from house to nearest safe water option (F)	W_F
Coping Capacity Component	W_R	Percentage of population aware of arsenic (G)	W_G
		Average distance from health center (H)	W_H
		Socio economic status (I)	W_I
Health Hazard Component	W_S	Percentage of arsenic patients (J)	W_J

IV. PROPOSED RESOURCE DISTRIBUTION METHOD USING ADSS

A. Decision for finding number of green tube wells required for a Mouza

ADSS follows tube well distribution proportionate to the Mouza VI or in other words according to the Arsenic crisis level whereas conventional distribution suggests distributing same number of tube wells for all Mouzas regardless their different level of severity. Once we have VI information, the system is now able to take decisions to find the number of red tube wells to be replaced in the selected Mouzas under a union in a thana. To calculate the new VI, overall hazard takes into account all the 10 parameters, whereas, arsenic hazard and health hazard consider only the arsenic and health related parameters. The formula for finding number of green tube wells required for a Mouza is:

$$RT[p] = \frac{VI[p]}{\sum_{i=1}^{m}(VI[i])} \times A$$

Where, m = Number of selected Mouzas for which the distribution is being considered,

p = The Mouza named "p" from m Mouzas,

RT[p] = Required number of tube wells for the Mouza "p",

VI[p] = Vulnerability Index of the Mouza named "p",

VI[i] = Vulnerability Index of the Mouza named "i",

A = Number of tube wells to be distributed set by the user.

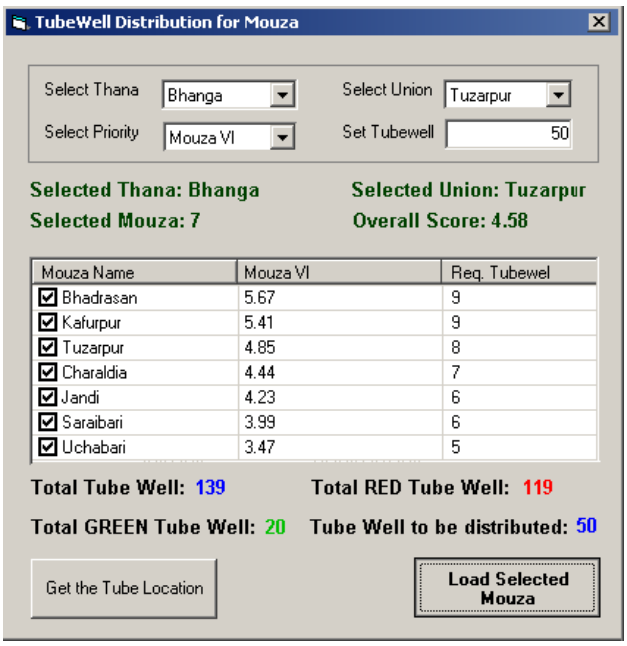


Fig. 3. User Interface for distributing Green tube wells based on Mouza VI

The user can select any one of three hazards such as overall hazard, arsenic hazard, and health hazard as the parameter to see the effect of changed VI. Selecting any of these priorities, the system returns the list of Mouzas along with their required tube wells.

Fig. 3 shows the list of Mouzas of "Tuzarpur" Union of "Bhanga" Thana, sorted by "Mouza VI" for distributing 50 green tube wells. Mouzas listed according to descending VI values, where greater Mouza VI means more affected Mouzas and requires more green tube wells. Arsenic Hazard includes only Arsenic related parameters (Percentage of red tube well, Average arsenic concentration, Percentage of population using red tube well).

Fig. 4 illustrates tube well distribution based on Arsenic Hazard. It shows the 7 Mouzas of "Tuzarpur" Union under

"Bhanga" Thana along with the required tube wells to be replaced. Since we are taking only 3 parameters out of 10 parameters, the distribution is different from the tube well distribution based on Mouza VI. For example, "Bhadrasan" Mouza is assigned 9 tube wells for the previous distribution but 10 tube wells are assigned for Arsenic Hazard distribution.

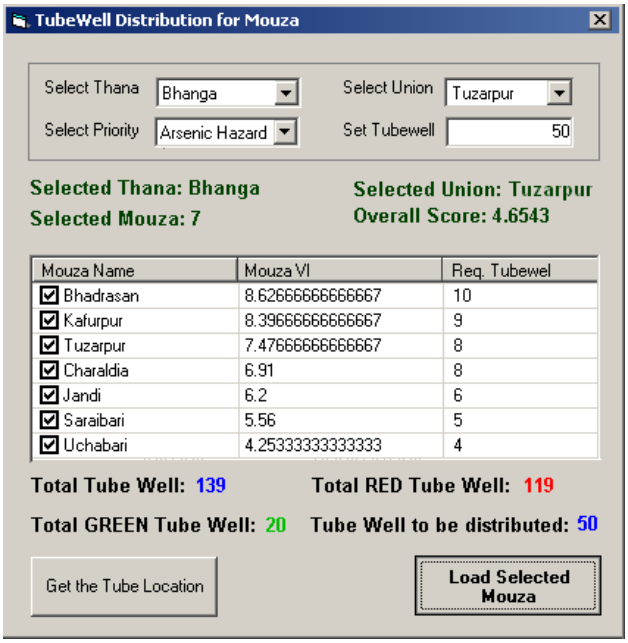


Fig. 4. User Interface for distributing Green tube wells based on Arsenic Hazard

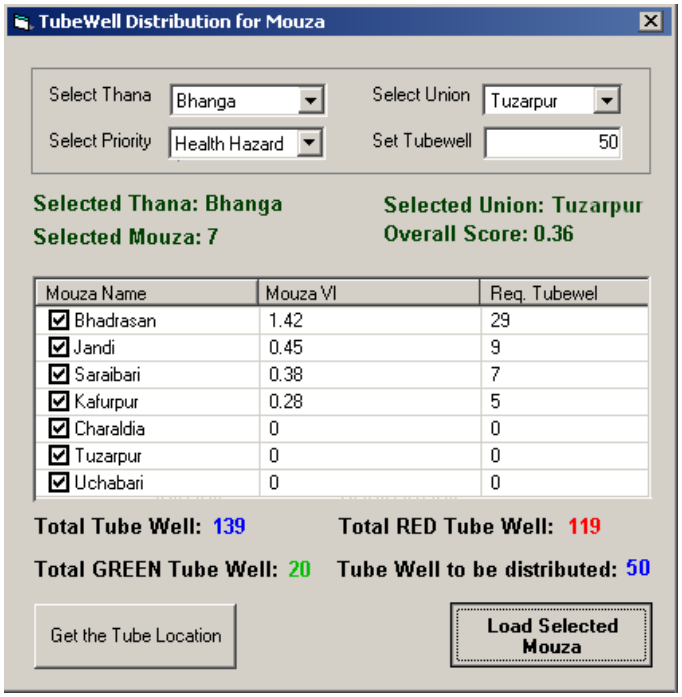


Fig. 5. User Interface for distributing Green tube wells based on Health Hazard

Fig. 5 illustrates tube well distribution based on Health Hazard. Since, we are taking only 1 parameter (Percentage of arsenic patients) out of 10 parameters, the distribution is different from the tube well distribution based on Mouza VI and Arsenic Hazard described previously. According to this distribution "Bhadarsan" Mouza is assigned 29 tube wells and since Charadia, Tuzarpur and Uchabari Mouza have no arsenic patients, they need no Tube wells.

B. Decision for finding suitable locations for replacing red tube wells

Once we have decisions for finding number of green tube wells required for a Mouza, we proceed to the second decision for finding suitable locations for replacing red tube wells. To start with, we first need to calculate a parameter called "tube well VI" for all red tube wells under a Mouza in order to be able to compare vulnerabilities among all tube wells. ADSS searches and takes all red tube well information, calculates tube well VI, based on different hazards and further selects the most vulnerable tube wells.

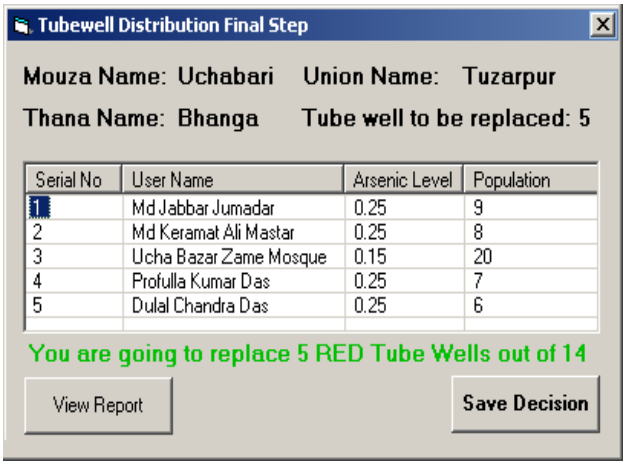


Fig. 6. User Interface for finding locations for replacing Red tube wells

Figure 6 shows worst 5 red tube wells out of 14 red tube wells in "Uchabari" Mouza under "Bhanga" Thana. It is to be noted that, both arsenic level value of red tube and number of people using that red tube well is considered in order to find vulnerable tube wells. For example, record no. 4 has greater arsenic value (0.25) than the 3rd record (0.15) but it is listed afterwards considering no. of people using red tube wells.

V. AN INTEGRATED SMART USER INTERFACE

An integrated smart user interface is designed for the users who will find an interactive, user-friendly, intelligible environment to work with and will work just like a "One Stop Solution", where users will find all of their related queries instantly without any hassle. The designed system is capable of comparing and differentiating between (Old VI and New VI) individually and also group-wise by a single click and can display data output in both tabular and spatial format.

As soon as we execute the program our proposed "Integrated Smart User Interface" will look like Fig. 7, where we need to set the parameters according to our requirements. There are 3 different tabs named "VI Parameters", "Simulation" and "Tabular and Spatial View". Simulation tab plays the vital role for the required outcome. After the user agrees to proceed, it will start working to update the database.

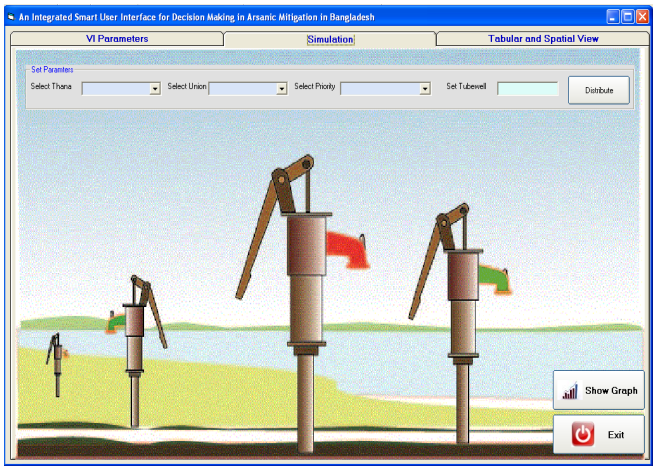


Fig. 7. An Integrated Smart User Interface

After VI calculation is done the appearance of the interface will look like as shown in Fig. 8. After calculating the new VI with given set of parameters or preferences, it will show the changes in terms of increase and decrease of VI in different Mouzas, average new VI per Mouza, average old VI per Mouza, and will list all the Mouzas with the new and old VI. The view option on the right enables the user to simulate the arsenic situation with different preferred output style.

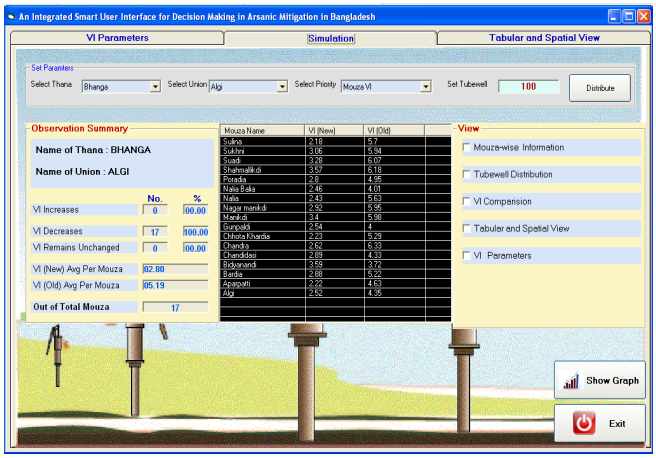


Fig. 8. After VI Calculation

Table II depicts the comparison between VI (New) and VI (Old) for a sample run. Here, in this table, we can see that there are 17 mouzas under Bhanga thana where for all the cases the values of VI (New) are decreasing. As we already discussed about the VI (Vulnerability Index) in the earlier section (section III) that the VI is nothing but a numerical value, which represents the arsenic situation of a Mouza comparing with other Mouzas in a thana or union. The greater the VI value, the greater the risk is. While ADSS analysis mainly focuses on finding the VI, which is calculated by some formulas based on mathematical, statistical and domain knowledge, is termed as VI (old), the new value that we see is VI (New) that have been calculated with the preferred parameters set by the user.

TABLE II. VI COMPARISON

Mouza Name	VI (New)	VI (Old)
Sulina	2.18	5.7
Sukhni	3.06	5.94
Suadi	3.28	6.07
Shahmallikdi	3.57	6.18
Poradia	2.8	4.95
Nalia Balia	2.46	4.01
Nalia	2.43	5.63
Nagar manikdi	2.92	5.95
Manikdi	3.4	5.98
Gunpaldi	2.54	4
Chhota Khardia	2.23	5.29
Chandra	2.62	6.33
Chandidasi	2.89	4.33
Bidyanandi	3.59	3.72
Bardia	2.88	5.22
Aparpatti	2.22	4.63
Algi	2.52	4.35

The "show graph" option in the interface displays the obtained simulated results graphically. A Graph has been shown in Fig. 9 based on the value of the Table II. It is clear from the Graph that, for all the cases the values of VI (New) are decreasing compared with VI (Old) after the distribution of the available resources in different Mouzas.

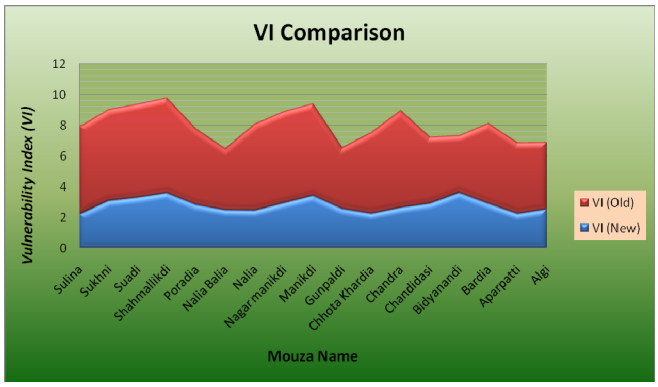


Fig. 9. VI Comparison

Our target is to display both the old and new VIs in the same graph to compare area-wise vulnerabilities. Once we calculate VI information, we can now show VI information by both in "Tabular and Spatial View" format. After the user select on the navigation area "Tabular and Spatial View", the

following figures (Fig. 10 and 11) will be displayed which is nothing but gives the user a both a tabular and pictorial overview about the severity of Arsenic effects in selected administrative boundary (e.g., Thana/ Mouza/ Union/ Upozila/ Village).

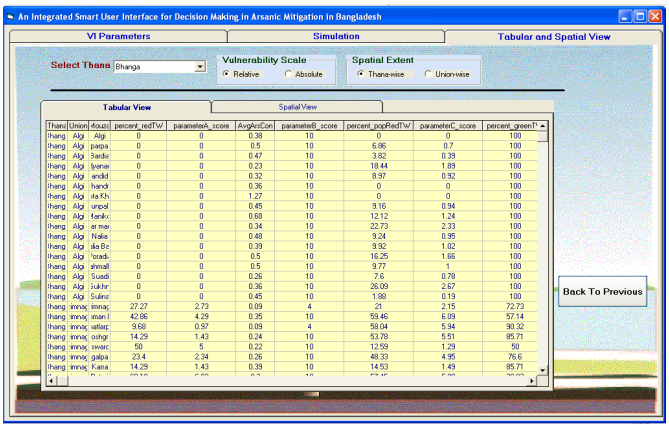


Fig. 10. VI Calculation (Tabular View)

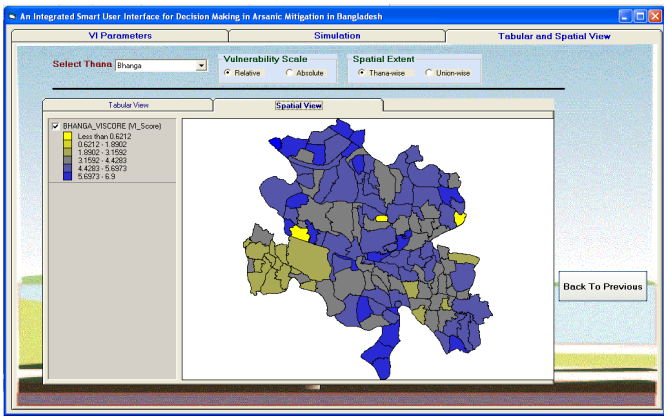


Fig. 11. VI Calculation (Spatial View)

VI. Conclusion

Arsenic contamination is an overwhelming problem in the context of Bangladesh. A number of studies were done previously to assess the situation and some database-oriented software was developed to measure the vulnerability in different areas. But not enough work is done on simulating the effect of resource allocation based on different parameters that affect arsenic vulnerability. In this paper, we have presented a decision support system for optimal allocation of resources such as tube wells in different areas. The proposed resource distribution framework along with the developed smart user interface is extensively useful for utilizing limited resources with high demand of people at arsenic contaminated area using optimized decision. It can be primarily used by the Government agencies for Arsenic Mitigation, as well as Research organizations working on Arsenic Mitigation in Bangladesh.

Acknowledgment

The data was collected from CEGIS (Center for Environment and Geographic Information System) who has conducted a pilot project for arsenic mitigation in Bangladesh covering three thanas namely Muradnagar of Comilla district, Sirajdikhan of Munshiganj district and Bhanga of Faridpur district from 2001 to 2005.

References

[1] Planning Commission, "Millennium Development Goals: Needs Assessment and Costing 2009-2015 Bangladesh", General Economics Division, Planning Commission, Government of the People's Republic of Bangladesh, Dhaka, Bangladesh, 2009.

[2] BBS, "Bangladesh Compendium of Environment Statistics", BBS, Dhaka, Bangladesh, 2004.

[3] UNICEF, "Arsenic Mitigation in Bangladesh", UNICEF-Bangladesh, Dhaka, Bangladesh, 2000.

[4] BBS & UNICEF, "Multiple Indicator Cluster Survey 2009", *BBS &* Ministry of Planning, Government of the People's Republic of Bangladesh & UNICEF, Dhaka, Bangladesh, 2010.

[5] Smedley, PL, Nicolli, HB, & Luo, ZD, "Arsenic in ground waters from major aquifers: sources, effects and potential mitigation", BGS Technical Report WC/99/38, BGS, NERC, Nottingham, UK, 2000.

[6] Matin, MA, Hassan, A, Sarke, MH, Elahi, MM, Shafey, MAL (February 2006). "CEGIS ADSS: Arsenic Decision Support System [online]. Avialable: http://www.cegisbd.com/soft13.htm

[7] Pidwirny, M., Geographic Information Systems (2011). Encyclopedia of Earth. Eds. Cutler J. Cleveland (Washington, D.C.: Environmental Information Coalition, National Council for Science and the Environment) [online]. Available: http://www.eoearth.org/article/Geographic_Information_Systems_%28GIS%29?topic=49501

[8] Druzdzel, MJ, and Flynn, RR., Decision Support Systems, Encyclopedia of Library and Information Science, 2nd Edition, Marcel Dekker Inc., New York, USA, 2002.

[9] Flanagan SV, and Zheng Y., Making Economics Sense for Arsenic Mitigation: A Case Study of Comilla District, Bangladesh, UNICEF Bangladesh, Dhaka, Bangladesh, 2011.

[10] Elahi, M.M., Amin, M.I., and Miah, M.R., "An optimal resource distribution method for arsenic mitigation in Bangladesh", *15th International Conference on Computer and Information Technology (ICCIT)*, 2012.

[11] http://www.cegisbd.com.

Proactive Approach of Making Eye Contact with the Target Human in Multi-Party Settings

Mohammed Moshiul Hoque[1], Dipankar Das[2]
[1]Dept. of Computer Science & Engineering
Chittagong University of Engineering & Technology,
Chittagong, Bangladesh
moshiul_240@cuet.ac.bd

Yoshinori Kobabashi[2], Yoshinori Kuno[2], Kaushik Deb[1]
[2]Dept. of Information & Computer Sciences
Saitama University
Saitama, Japan
{dipankar, yosinori, kuno}cv.ics.saitama-u.ac.jp

***Abstract*— Establishing eye contact with the target agent plays a central role both in human-human and human-robot communications. However, it is not so easy task for the robot to meet eye contact with a particular human. Especially, when the robot and the target human are facing each other initially or the target human is intensely involved his/her task. In this paper, we proposed a conceptual model of eye contact for social robots consisting of three parts: attracting attention, establishing gaze crossing, and displaying gaze awareness. Evaluation experiments with 36 participants reveal the effectiveness of the proposed model in three viewing situations (near peripheral field of view, far peripheral field of view, and out of field of view).**

***Keywords*— Human-robot interaction, eye contact, gaze crossing, and gaze awareness.**

I. Introduction

Currently work in robotics is expanding from industrial robots to robots that are employed in the living environment. Human-robot interaction (HRI) is an interdisciplinary research field aimed at improving the interaction between human beings and robots and to develop robots that are capable of functioning effectively in real-world domains, working and collaborating with humans in their daily activities. To cope with the human world, robots should be capable to perform a plenty of social and cognitive behaviors as human being does with other human. Although a number of significant challenges remained unsolved related to the social capabilities of robots, the robot that can pro-actively meets eye contact with human is also an important research issue in the realm of natural HRI.

Eye contact is a phenomenon that occurs when two people cross their gaze (i.e., looking at each other). Meeting eye contact is considered as an important prerequisite to start a natural interaction process between two partners [1, 2]. Moreover, it plays several important roles in face-to-face communication, such as, exhibiting one's attention [3], regulating turn taking [3], and joint visual attention [4]. However, it is difficult to set up such eye contact for one person while the target person is not facing toward him/her or while s/he is intensely attending at his/her work. In that case, robots should use a proactive approach to the intended human for making eye contact.

In this paper, we proposed a conceptual model of eye contact that can pro-actively makes eye contact with the target participant from multiple participants. This model consists of three parts: attracting attention, establishing gaze crossing, and displaying gaze awareness. It is readily apparent that a robot can attract a human attention utilizing voice or sound cues. But vocal utterances are also certain to attract the attention of others who are not intended targets. However, we would like the robot to be able to selectively attract only the target participant while avoiding attracting other participant's attention as much as possible. Thus, the robot should try with the non-verbal action first. According to James [5], object motion draw people's attention. Therefore, we are decided to use robot's head motions to attract the target participant's attention. However, it apparent that visual stimuli offered by the robot's nonverbal behaviors cannot affect a participant if s/he is in a position from which s/he cannot see the robot. In this situation, the use of touch or sound cues should be considered a last resort.

Attention attraction can produce observable behavioral responses such as eye, or head movements. Therefore, if the target person is attracted by the robot behaviors, a person will turn toward the robot. After facing the robot, robot should interpret human looking response and adjust its head/body orientation to ensure face-to-face orientation. After capturing the attention of the intended recipient, the robot needs to make the person notice clearly that it is looking at none other than him/her after interpreting looking response. To solve this problem the robot should be able to display its awareness explicitly by some actions.

II. Related Work

The capability of robots that can control human attention is in a rudimentary stage still now. Several previous HRI studies have addressed greeting behavior to initiates human-robot conversation at a social distance [6, 7]. Some robots were equipped with the capability to encourage people to initiate interaction by some non-verbal cues such as body orientation and gaze [8], approaching path [9], standing position [10], following behaviors [11]. These robots also wait for people to approach them. Several others robotic systems were developed to establish eye-contact [12, 13]. These robots are supposed to make eye contact with humans by gaze crossing (i.e., turning their eyes [cameras] toward the human faces). Most of these

studies assumed that the target person faces the robot and intends to talk with it; however, in reality this assumption may not hold. Although such a passive attitude can work in some situations, many situations require a robot to use an active approach [14, 15]. Moreover, psychological studies show that gaze crossing function alone may not be enough to make eye contact [16]. Yoshika et al. [17] also mentioned that simply staring is not always sufficient for a robot to make someone feel that they are being looked at.

Certainly, a robot that approaches humans and starts an interaction pro-actively for making eye contact should be perceived to act more natural than a passive machine. Some robots were equipped with the capability to initiate interaction pro-actively with humans by non-verbal behaviors such as predicting the trajectories [18], path position [19] and body orientation [20]. These systems did not care about the other people's attention interruption and it is certain that all people may be attracted and looking toward the robot when it is moving or walking. After crossing the gaze with the intended recipient, the robot needs to make the person notice clearly that it is looking at none other than him/her after interpreting looking response. To solve this problem the robot should be able to display its awareness explicitly by some actions. Some robotic systems were incorporates gaze awareness functions by facial expression (i.e., smiling) [21], head movements [22], and ear blinks [23]. In order to produce smiling expression, they used a flat screen monitor as the robot's head. A flat screen is unnatural as a face. Moreover, these systems are mechanically very complex and as such expensive to design, construct and maintain. Our previous studies reveals that a robot's head motions actions can attract human attention and robot can display its gaze awareness functions by blinking its eyes. However, we dealt only with cases involving a single participant. Therefore, how the robot should behave to establish the eye contact with the target human alone in multi-party setting is unexplored.

Although these studies provide some evidences that the robot can attracts and meet people's attention in different viewing situations, a systematic study of how the robot pro-actively establish eye contact with the target human alone in multi-party setting while s/he engaged her/his work are still lacking. The following questions remain unanswered; Can robot can attract the target human's attention in different viewing situations? Can a robot behave in such a way by which only the target person feels eye contact from multiple persons? Can a robot convey its gaze awareness behaviors explicitly to the target person?

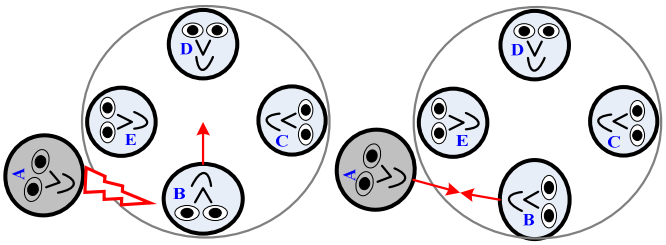
III. Establishing Eye Contact in HRI

In this section, we operationalize the eye contact process. We also describe how this process can apply in an HRI platform to make eye contact with the target human.

A. An Eye Contact Episode

A canonical eye contact episode, between the two agents (i.e., the robot and the target participant (TP)), can be described in three major steps: attracting attention, establishing gaze crossing, and displaying gaze awareness. Fig. 1 shows a simple depiction of the eye contact process.

- **Attracting attention:** Attention is the cognitive process of selectively concentrating on one aspect of the environment while ignoring other things. Attracting attention refers to one's ability to capture another's attention, shifting it from its existing direction towards him/her for establishing a communication channel through eye contact. Fig. 1 (a) shows that agent A is sending an action to attract the agent B's attention. Although eyes are the primary source of information about the direction of attention, there are many situations where determining the viewer's eye direction is impossible or infeasible. Especially when people are further apart or are not facing each other, head orientation becomes a stronger cue than information from the eyes in determining direction of attention [24]. In our eye contact process, we employ head direction to monitor the focus of the participants' attention. The robot apply the actions (i.e., head turning, head shaking, and reference terms) based on the attentional focus of the TP.



(a) Attracting attention (b) Establishing gaze crossing

Figure 1. Eye contact episode. Direction of arrows indicates the agent's attentional focus.

- **Establishing gaze crossing:** Gaze crossing between two people refers to the looking at each other eyes or face thus acting simultaneously as sender and recipient [25]. Fig. 1 (b) illustrates the gaze crossing behaviors between agents A and B. Attracting the attention of the TP can produce observable behavioral responses such as eye or head movements, or shifts in body orientation. Therefore, if the TP is attracted by the robot's actions, s/he will turn toward it, which will make establishing gaze crossing.
- **Displaying gaze awareness:** refers to display the awareness of sender against to the looking response intended recipient. After crossing the gaze with the TP, the robot needs to make the person notice clearly that it is looking at none other than him/her. Gaze awareness behavior of the robot will helps to create the feeling of being looked at of the TP. Thus, to be most effective the robot should also display gaze awareness explicitly through some actions (e.g., facial expressions or nodding) after gaze crossing.

B. Proposed Approach

Humans usually first turn their head toward the person with whom they would like to communicate. If the target human does not respond, s/he will attempt to gain their attention with stronger signals (e.g., waving a hand, shaking the head, utilizing sound cues or even using their voice). In a

natural HRI interaction scenario, robots should utilize the same conventions as humans. In other words, the robot should start with a weak action to avoid attracting the attention of individuals other than the target person, and then use stronger actions if its initial attempt fails to attract his/her attention. Based on surveying previous literature [12], we decided to have our robot turn its head (to look towards the person) as the weakest action. If the robot does not attract the attention of the target agent through this head turn, we decided that it would next employ head shaking and then uttering reference terms in this order as a stronger action.

An attention control event is executed by a finite-state-machine model as shown in Fig. 2. To initiate the attention control, the robot begins to observe the current direction of the agent's attention by tracking his/her head. After recognizing the viewing situation of the target agent (TA), the robot usually turns its head first toward the TA, and commences shaking its head, and then uttering reference terms (if necessary) to attract his/her attention. For the head turning action, we adjusted the pan speed of the pan-tilt unit at 120^0/second. For the head shaking action, the robot shook its head back and forth ($\pm 30^0$) from its initial position. This meant that the robot turned its head 30^0 left and 30^0 right. The head-shaking speed was adjusted at 240^0/second. We would like to use the words, 'excuse me' as the reference terms. However, the robot waits about 2 seconds after each attempt for the TA to respond by looking in its direction.

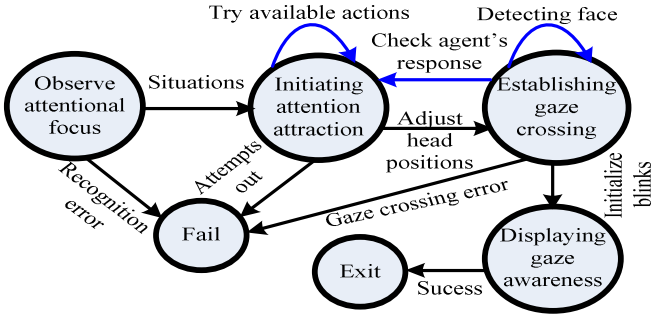


Fig. 2. An integrated behavioral model of eye contact.

If the robot is successful in attracting the TA's attention, the two agents will experience gaze crossing. Thus, the robot considers the TA to have responded to its actions if he/she looks at the robot within the expected time-frame. It is able to recognize whether this is so by detecting the front of his/her face in the camera image. The robot performs a blinking action to display gaze awareness. A previous study has shown that blinking is one of the most important cues for forming a person's impressions [26]. The robot's eye blinks are produced by the rapid closing and opening of the eyelids depicted in the CG images, and displayed through the LED projector on the robot's eyes. The TA feels that he/she has made eye contact with the robot, and comes to expect the initiation of an attention shift phase. The robot meets the TA's gaze, and blinks its eyes in response to only them in the event that both target and non-target agents are looking at it simultaneously. The robot considers the case to be a failure if it cannot establish the eye contact with the TA within 4 seconds. Psychological studies have shown that silences of more than 4 seconds become embarrassing because they imply a break in the thread of communication [27].

IV. System Overview

For the HRI experiments we developed a robotic head as shown in Fig. 3. This figure shows a prototype robotic head and the corresponding outputs of its several software modules. The head consists of a spherical 3D mask, an LED projector (3M pocket projector, MPro150), a laser range sensor (URG-04LX by Hokuyo Electric Machinery), three USB cameras (Logicool Inc., Qcam), and a pan-tilt unit (Directed Perception Inc., PTU-D46). The LED projector projects CG-generated eyes on the mask. In the current implementation, one USB camera is attached to the robot's head (as shown in Fig. 3). The other two cameras and the laser sensor are affixed to the tripods, placed at an appropriate position for observing participants' heads as well as their bodies.

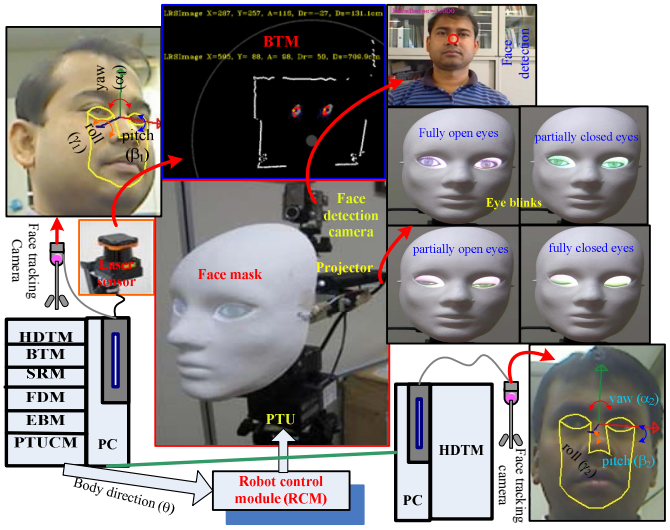


Fig. 3. System overview consists of several software modules including HDTM, BTM, SRM, FDM, EBM, PTUCM, and RCM.

The system utilizes two general purpose computers (Windows XP), each connected to a USB camera. One PC is installed with the head detection and tracking modules (HDTM) for tracking the head of the non-target participant. To detect, track, and compute the direction of the participant's head in real time (30 frame/sec), we use FaceAPI [28] by Seeing Machines Inc. It can measure 3D head direction within a 30 error at 30 frame/second. The other PC is installed with three software modules, namely HDTM, the body tracking module (BTM), and the robot control module (RCM). This PC detects and tracks the bodies of multiple people in the range sensor data using the BTM. The human body model is consequently represented with the center coordinates of an ellipse [x, y] and its principle axis direction (θ). These parameters are estimated in each frame by a particle filter framework. BTM can locate a human body within errors of 6cm for position, and 6^0 for orientation. A detailed description of the BTM is provided in [29]. Both PC's are connected through a wired network. The RCM consists of four sub modules: a situation recognition module (SRM), face-detection module (FDM), eye-blinking module (EBM), and

pan-tilt unit control module (PTUCM). The RCM integrates all sensor processing results of both computers.

To assess the current situation (i.e., where the human is currently looking), one observes the head direction estimated by the HDTM. From the results of the HDTM, the SRM recognizes the existing viewing situation (with 99.4% accuracy) and the direction of the relevant object in terms of yaw (α), and pitch (β) movements of the head by using a set of predefined rules. For each rule, we set the values of (α), and (β) by observing several experimental trials. For example, *if the current head direction (of the human with respect to the robot) is within* $-10^0 \leq \alpha \leq +10^0$ *and* $-10^0 \leq \beta \leq +10^0$ *and remains in the same direction for 30 frames, system recognizes the situation as the central field of view*. The results of the SRM are sent to the PTUCM to initiate the attention control process, and the robot turns toward the human based on the results provided by the BTM. The robot considers that the participant has responded successfully to its actions if the TA looks at the robot within the expected time-frame. If this step is successful, the FDM uses the image from the forehead camera to detect the front of the target participant's face [30]. After detecting the target agent's face, the FDM sends the results to the EBM for exhibiting eye blinks to let the human know that the robot is aware of his/her gaze. The EBM generates the eye blinks at a rate of 1 blink/second up to three times [14]. All the robot's head actions are performed by the PTU, with the actual control signal coming from several modules.

V. Effects of Attention Attraction Behaviors

We conducted an experiment to verify the effectiveness of proposed eye contact process for attracting the target participant's attention in different viewing situations. In these situations, we assumed the target participant is being engaged in such a work that did not absorb much attention to perform.

A. Participants

A total of 36 graduate students from Saitama University participated in the experiment. Participants are randomly assigned into three groups according to the viewing situations (NPFOV, FPFOV, and OFOV). In each group, two participants interacted at a time, one of them interacting with each robot as a target and another as a non-target. Each pairs of participants interacted in two sessions and the role of them were exchanged in each session.

B. Design and Procedures

As a low attention-absorption task, we considered a scenario: 'watching the paintings'. To prompt participants to look in various directions, we hung seven paintings (P1-P7) on the wall at the same height (just above the eye level of the participants). These paintings were placed in such a way that, when observed from a participant's sitting position, they covered their whole field of view (close to 180^0). Two robots were placed in the participant's left and right monocular fields of view. One robot (SR) was stationary all times, while the other (ECR) performed the eye contact process. Fig. 4 (a) shows the schematic setting of the experiment.

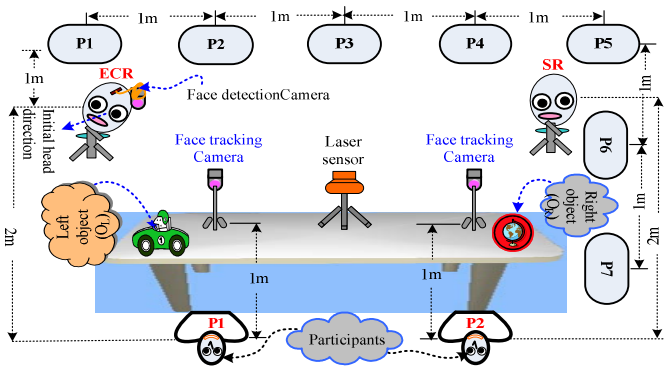


Fig. 4. Schematic setting of the experiment where ECR means the eye contact robot, SR means the static robot, P1 and P2 represents the participant 1 and 2.

Our intention was to let the participants evaluate the various behaviors of the robot as it attempted to acquire their attention when they were not initially looking in its direction. For this purpose, a pair of participants were asked to sit down on chairs and asked to look around at the paintings. Since the positions of the robots were fixed, the participants detected them in their various fields of view as they moved their heads and bodies around. Our proposed robot randomly selected one of the participants as a target and the other as a non-target. The participants took part in two trials. The robot tried to attract the target participant's attention while he/she was looking at different paintings so that it could obtain data for two types of viewing situation: the near peripheral field of view (e.g., looking at pictures P3, P4 or P5), and the far peripheral field of view (e.g., looking at pictures P6 or P7). In this experiment, if the target participants did not gaze at the robot within the expected time frame following the robot's actions, then the robot considered the case to be a failure and completed the experiment. Fig. 5 depicts a scene from the experiment. Two video cameras were placed in appropriate positions to capture all interactions during the experiment.

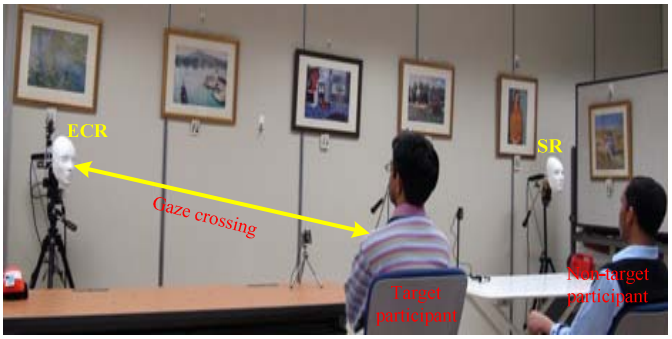


Fig. 5. The ECR and the target participant meeting the eye contact (right).

C. Design and Procedures

To evaluate the effectiveness of our proposed method, we compared it with two alternative methods. Each participant experienced the following three methods one after another.

- **HT-robot (HTR)**: The robot always applies head turning action whatever the situation is to attract participant's attention.
- **HS-robot (HSR)**: The robot always applies head shaking action in all situations to attract human's attention.

- **Eye-contact robot (ECR)**: This is our proposed robot. It applies head turn for NPFOV, head shaking for FPFOV, and reference terms for OFOV respectively.

D. Results

The experiment had a between-subject design. 12 participants experienced as the target in each conditions. Thus, we observed a total of 108 (12 [participants] × 3 [situations] × 3 [methods]) interaction videos for all robots. For overall evaluation, we measured the one qualitative item such as 'success ratio' by observing the experimental videos. This was measured by the ratio of the number of target participants looking at the robot (N_{TL}) and the number of times that the robot attempted to attract their attention (N_{AA}). Table I summarizes the quantitative measures in attracting participant's attention by robots in three viewing situations.

TABLE I. SUCCESS RATIO IN DIFFERENT VIEWING SITUATIONS

Methods	N_{TL}/N_{AA} (%)			Overall success ratio
	NPFOV	*FPFOV*	*OFOV*	
HTR	11/12 (92%)	01/12 (8.3%)	01/12 (8.3%)	36.2%
HSR	12/12 (100%)	10/12 (83.3%)	02/12 (16.6%)	66.7%
ECR	11/12 (92%)	11/12 (92%)	10/12 (83.3%)	88.9%

Results show that most of the participants responded to HTR in NPFOV. However, in the case of FPFOV and OFOV, only 8% was attracted by HTR. HSR were successful in attracting most of the participant's attention both in NPFOV (100%) and FPFOV (83%) situations except OFOV (16%) situation. On the other hand, ECR was successful in attracting participant's attention in all cases (92% for NPFOV and FPFOV situations, and 83% for OFOV situation). We conducted a repeated measures of analysis of variance (ANOVA) for NPFOV situation, which does not show any significant differences among the three robots ($F(2, 22) = 0.48$, $p = 0.62$, $\eta^2 = 0.02$) . Thus, we seen that the robot no need to shaking its head or uttering any reference terms to attract target person's attention toward it, tuning head toward the target person action is enough when s/he sensed the robot's action in his/her NPFOV situation.

We also conducted an ANOVA analysis for FPFOV situation, which shows significant differences among robots ($F(2, 22) = 24.41$, $p<0.001$, $\eta^2 = 0.59$). Multiple comparisons with post-hoc Tukey test show significant differences between both pairs (HTR vs. HSR: $p < 0.001$, and HTR vs. ECR: $p < 0.0001$). No significant differences are found between HSR and ECR robots in attracting participant's attention ($p = 0.58$). Therefore, we can say that actor's basic action (i.e, head turning) did not affected the current attentional focus of the target person and hence the robot should be applied stronger action (head shaking in our study) when s/he sensed the actor's action from FPFOV.

There is a significant differences were found for OFOV situation in three conditions (ANOVA: $F(2, 22) = 17.08$, $p<0.001$, $\eta^2 = 0.48$). Multiple comparisons with post-hoc Tukey test show that ECR is significantly differ from HTR ($p<0.0001$) and HSR ($p=0.0007$). However, no significant difference is found between HTR and HSR ($p = 0.58$). These results revealed that any movements of the robot's did not affected the target person's existing focus of attention because that motions cannot be perceived by his/her eyes. That means, in OFOV situation, any kind of motions did not effective, and in that case, only voice signal can be capture humans' attention effectively.

VI. EFFECTS OF EYE BLINKS

The purpose of this experiment is to verify the effects of eye blinking action for making eye contact with the target participant.

A. Methods

18 graduate students (average age = 26.8) at Saitama University participated in the experiment. The experiment environment was almost the same as previous except the behaviors of the robot ECR. It adjust its head position (if needed) to cross the gaze of the attracted target participant. After gaze crossing, the robot displays its eye blinking actions to the attracted target participant as gaze awareness. However, it did not perform any blinking action if the non-target participant is looked at it before the target participant.

B. Experimental Conditions

The experiment had a within-subject design, and the order of all experimental trials was counterbalanced. Every participant experienced all two conditions. To verify the effect of eye blinks, we prepared the following two conditions:

- **Eye contact robot with no blinks (ECR+NB)**: The robot uses attention attraction behaviors as proposed in section III- B. If the target participant is looked at it, the robot recognizes his/her face, waits about three seconds, and then turns its head toward another direction.
- **Eye contact robot with blinks (ECR+B)**: This is our proposed robot. The robot recognizes his/her face while s/he is looking at it. After detecting the face of the target participant, the robot starts blinking its eyes about three times (1 blink per second), then turns its head toward another direction. Detail description about the behavior of this robot is given section III-B.

C. Results

After interacting with the robot in both conditions, the participants were asked to provide evaluations on a 1-to-7 point Likert scale in the questionnaire, where 1 stands for the lowest and 7 for the highest. The questionnaire had the following items:

- **Attention attraction**: Did you feel that the robot attracted your attention?
- **The feeling of being making eye contact**: Did behaviors of the robot created your feeling of making eye contact?

Fig. 6 shows the participants response on each question. A repeated measure of analysis of variance (ANOVA) attention attraction does not shows any significant difference between the robots ($F(1,35) = 0.32$, $p = 0.57$, $\eta^2 = 0.56$).

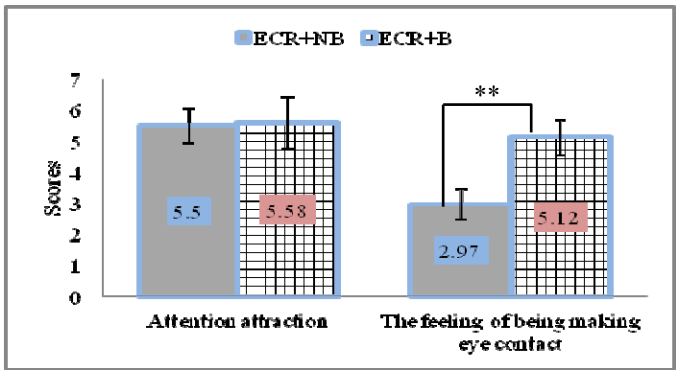


Fig. 6. Subjective evaluation results. Error bars indicates the standard deviation and ** means significant differences.

VII. Conclusion

The primary focus of our work is to develop a framework for robots that would make eye contact with human when it intended. For serving our purpose, we have proposed a robotic framework that can be made eye contact with the target participant in three stages: attracting attention, establishing gaze crossing, and displaying gaze awareness. The success of a particular action of a robot depends on several factors, such as distance between a robot and a target human, spatial arrangement between them, duration of an action, direction of attention of the target participant, level of attention of the target participant the task involved, and so on. In this study, we have shown that robot's actions are effective to gain the target participants attention in three viewing situations. We have shown that eye blinking in response to the human is effective to make a feel that the robot can understand his/her response and succeeded in giving the target participants a stronger feeling of being making eye contact. We have considered a particular scenario in this paper. If a person is intensely paying attention to his/her works or talking with another person, the robot needs to adapt other strategies. Moreover, this work did not consider the behaviors of the non-target participants and did not measure the total time needed for making eye contact in different situations. These issues are left for future work.

References

[1] M. Argyle, Bodily Communication, Routledge, London, 1988.
[2] T. Striano, V. M. Reid, and S. Hoeh, "Neural mechanisms of joint attention in infancy", Euro. J. Neurosci., vol. 23, no. 10, pp. 2819-2823, 2006.
[3] C. Kleinke, "Gaze and eye contact: A research review", psychological bulletin", vol. 100, no. 1, pp. 78-100, 1985.
[4] T. Farroni, E. M. Mansfield, C. Lai, and M. H. Johnson, "Infant perceiving and acting on the eyes: Tests of an evolutionary hypothesis", J. of Exp. Child Psy., vol. 85, pp. 199-212, 2003.
[5] W. James, The Principles of Psychology, vol. 2, Dover, New York, Dover, 1950.
[6] K. Hayashi et al., "Humanoid robots as a passive-social medium-A field experiment at a train station", in HRI, pp. 137-144, 2007.
[7] M. Shiomi, T. Kanda, H. Ishiguro, and N. Hagita, "A larger audience, Please! Encouraging people to listen to a guide robot", in International Conference on HRI, pp. 31-38, 2010.
[8] H. Kuzuoka, Y. Suzuki, J. Yamashita, and K. Yamazaki, "Reconfiguring spatial formation arrangement by robot body orientation", in Int. Con. on HRI, pp. 285-292, 2010.
[9] S. Satake, T. Kanda, D. F. Glas, M. Imai, H. Ishiguro, and N. Hagita, "How to approach humans? Strategies for social robots to initiate interaction", in Int. Con. on HRI, pp. 109-116, 2009.
[10] F. Yamaoka, T. Kanda, H. Ishiguro, and N. Hagita, "A model of proximity control for information presenting robot", IEEE Trans. on Robot., vol. 26, pp. 187-195, 2010.
[11] R. Gockley, J. Forlizzi, and R. Simmons, "Natural person following behavior for social robots", in HRI, pp. 17-24, 2007.
[12] B. Mutlu, T. Shiwa, T. Kanda, H. Ishiguro, N. Hagita, "Footing in human robot conversations: How robots might shape participant roles using gaze cues", in HRI, pp. 61-68, 2009.
[13] C. Breazeal, "Toward sociable robots", J. of Robo. & Autono. System, vol. 42, pp. 162, 2003.
[14] M. M. Hoque, T. Onoki, Y. Kobayashi, and Y. Kuno, "Effect of robot's gaze behaviors for attracting and controlling human attention", J. Adv. Robo., vol. 27, no. 11, pp. 813-829, 2013.
[15] C. Shi, M. Shimada, T. Kanda, H. Ishiguro, and N. Hagita, "Spatial formation model for initiating conversation", in International Conference on RSS, 2011.
[16] M. V. Cranch. The Role of Orienting Behavior in Human Interaction, Plenum Press, New York, 1971.
[17] Y. Yoshikawa et al., "The effects of responsive eye movement and blinking behavior in a communication robot", in International Conference on IROS, pp. 4564-4569, 2006.
[18] S. Satake et al., "How to approach humans? Strategies for social robots to initiate interaction", in HRI, pp. 109-116, 2009.
[19] M. Buss et al., "Towards proactive human-robot interaction in human environments", in CogInfocom, pp. 1-6, 2011.
[20] E. A. Sisbot, L. F. Marin-urias, R. Alami, and T. Simeon, "A human aware mobile robot motion planner", IEEE Trans. of Robotics, vol. 23, no. 5, pp. 874-883, 2007.
[21] D. Miyauchi, A. Nakamura, Y. Kuno, "Bidirectional eye contact for human-robot communication", IEICE Trans. of Info. & System, vol. 88-D, vol. 11, pp. 2509, 2005.
[22] Y. Yoshikawa, K. Shinozawa, H. Ishiguro, N. Hagita, and T. Miyamoto, "Responsive robot gaze to interaction partner", in Internal Conference on RSS, 2006.
[23] C. M. Huang, A. L. Thomaz, "Effects of responding to, initiating and ensuring joint attention in human-robot interaction", in International Conference on Ro-man, pp. 65-71, 2011.
[24] N. J. Emery, "The eyes have it: The neuroethology, function and evolution of social gaze", Neurosci. & Biobeha. Rev., vol. 24, no. 6, pp. 581-604, 2000.
[25] M. V. Cranach, J. H. Ellgring, "The perception of looking Behavior", Academic Press, London, 1973.
[26] K. Takashima, Y. Omori, Y. Yoshimoto, Y. Itoh, Y. Kitamura, F. Kishino, "Effects of Avater's blinking animation on person impressions", in Int. Conf. on GI, pp. 169-176, 2008.
[27] R. B. Miller, "Response time in man-computer conversational transactions", in AFIPS Fall Joint Computer Conference, pp. 267-277, 1968.
[28] FaceAPI: Face Tracking for OEM Product Development, Seeing Machines Int., http://www.faceapi.com, 2010.
[29] Y. Kobayashi, and Y. Kuno, "People tracking using integrated sensors for human robot interaction", in IT, pp. 1597-1602, 2010.
[30] G. Bradsky, A. Kaehler, V. Pisarevskky, "Learning based computer vision with Intel's open source computer vision library", J. Intel Tech., vol. 9, pp. 119-130, 2005.

Design and Implementation of an FPGA-based Real-time Face Recognition System using Spatial Correlation Function

Mohammad Shamim Imtiaz and Khan Wahid
Department of Electrical and Computer Engineering
University of Saskatchewan
Saskatoon, SK, Canada
msi756@mail.usask.ca and khan.wahid@usask.ca

Shaikh Anowarul Fattah
Department of Electrical and Electronic Engineering
Bangladesh University of Engineering and Technology
Dhaka-1000, Bangladesh
sfattah@princeton.edu

Abstract—**This paper presents a simple and efficient design of a face recognition system, where feature extraction algorithm is employed based on the principle of spatial cross-correlation. In the feature extraction process, instead of processing the entire image at a time, only a pair of rows or columns of an image is considered which makes the algorithm very efficient and low-cost. Considering the cross-correlations between the pairs, a unique 1-D signature of a 2-D face image is obtained which represents the variation in the face geometry along the vertical or horizontal direction. It is shown that the resulting vertical and horizontal features provide high compactness within the class and separation between the classes. It is found that the proposed design can provide a satisfactory recognition performance for different standard databases. The results of hardware implementation in terms of resources used and processing speed have also been presented.**

***Keywords*—Cross-Correlation, Spatial feature extraction, Face geometry, Classification, Face recognition, Verilog, FPGA**

I. INTRODUCTION

FACIAL appearance serves as a natural and effective means for recognition of a person by another. There are two main biometric techniques; Intrusive and nonintrusive. Face recognition system that employs a nonintrusive biometric approach is increasingly in demand for defense, security and commercial applications, as it protects both safety and privacy during the process [1]-[3]. Such systems generate a lot of raw data in real-time. A significant amount of processing power is a prerequisite for this sort of applications. In addition, complex algorithms for feature extraction of images make the hardware more complex and expensive. Several methods have been proposed for the implementation of face recognition system such as elastic subspace analysis method [4]-[5], Linear discriminant analysis (LDA), 3D modeling [7], Hidden Markov models [8], Bayesian analysis [9] and elastic graph matching [10].

Face Recognition is a high dimensional pattern recognition problem and the intrinsic dimensionality of the face space is much lower than the dimensionality of the raw image space. This fact is the starting point of the use of subspace methods such as Principal Component Analysis and Independent Component Analysis which are considered to be very successful methods among face recognition algorithms. However, they suffer from two limitations. First, the discriminative power of these methods decreases when the size of the training images increases [11]. Second, the computational load and memory requirement for calculating eigenvectors increase dramatically for large databases [11]. Although elastic graph matching is a precise face recognition approach, its computation is time consuming; it usually cannot satisfy the practical demands [10]. In 3D modeling approach, the existing sensors are not ideally suited for 3D based person identification as they are not dynamic, and often cannot provide reliable data from incompliant subjects, wearing glasses, in arbitrary conditions. Although 3D data is not subject to changes in illuminations, it is affected by other artifacts such as changes in expression, and holes in data caused by imaging effects [7]. Linear Discriminant Analysis (LDA) has been widely used for face recognition [12]. However, being a linear technique, it may not perform well when severe non-linearity is involved. Hidden Markov model based face recognition system leads to a large dimension of the observation vectors and it leads to high computational complexity of the training and detection/recognition systems [8]. Bayesian subspace analysis has been successfully applied in face recognition. However, the direct application of the algorithm is much more computationally intensive [9]. All these methods for face recognition have impressive recognition accuracy with computationally complex hardware design. Still researchers are looking forward o design a system with high recognition accuracy with a cost effective hardware design. This is the main motivation behind our current research work.

In this paper, we intend to adopt the holistic approach for precisely capturing the variations in the whole image. In order to capture variations in face geometry along the vertical and horizontal direction, cross-correlation operation is performed considering pairs of consecutive rows and column of an image data. A simple minimum distance based Euclidean classifier is used for template matching process. The proposed feature extraction method is computationally efficient and cost effective for real life implementation.

II. Derivation of Proposed Algorithm and Classifier

A. Cross-correlation based signature:

Face geometry varies in different region of a human face. Compared to the entire face image, certain portions such as eyes, nose and lips carry more information. These high informative zones can be used to identify a person due to the existence of low similarity in these zones compared to other zones [13]. Hence the idea introduced here is to exploit the discriminative features of the high informative zones of a face image for pattern classification. One effective way to measure the similarity of two vectors is the Cross-correlation. For given two finite dimensional real sequences $\mathrm{x(n)} = 0,1,2,3,\ldots,\mathrm{N}$ and $\mathrm{y(n)} = 0,1,2,3,\ldots,\mathrm{M}$ with $\mathrm{N} > M$, the $\mathrm{m^{th}}$ lag of cross correlation can be estimated as

$$p_{xy}(m) = \frac{1}{N}\sum_{n=0}^{N-1-|m|} x(n)y(n+|m|) \qquad (1)$$

where, $m \geq 0$. As images can be treated as 2-D array of pixel intensity values, each row and column can be treated separately as a sequence. In order to measure the similarity in a portion of the image, one may compute cross-correlation between two rows of the image using equation (1) considering $\mathrm{x(n)}$ and $\mathrm{y(n)}$ as the array of pixel intensities. One simple way to comment on the similarity of two sequences is to consider the zero lag value of the cross-correlation, i.e., $\mathrm{p_{xy}(0)}$. Use of lags other than the zero lag may be avoided as it unnecessary increases the computational burden without providing any significant benefit in improving feature quality [14]. If two rows are considered for cross-correlation and the similarity between them is very low (i.e. rows from forehead), the cross-correlation coefficient at the zero lag would be high in magnitude and vice versa. Thus, the value of the cross-correlation coefficient at the zero lag between two rows at the zero lags varies in accordance with the degree of similarity between these two rows. Considering the zero lag cross-correlation values of consecutive pair of rows, the vertical feature of a face image is extracted and similarly the horizontal feature is extracted considering the consecutive pair of columns. Concatenation of these two features provides the proposed concatenated feature vector, which is ultimately used for face recognition. Only matrix transpose, multiplier, sequential adder and divider would be enough to design feature extraction algorithm. As we considered matrix as a whole we need matrix transpose module.

B. Classifier:

In order to test the recognition accuracy by using proposed extracted feature, different classifiers can be used. However, in view of keeping the overall computational complexity very low, a distance-based classifier is employed. In the training phase, using the concatenated feature the feature template is generated. It is obtained by using the zero-lag value of the cross-correlation function. Then the template matching process is carried out based on distance-based Euclidean classifier. If the feature vector of the test image of a person is $F_{test}(l)$ and the feature vector of mth person in the nth pose from the training image database is $F^{m,n}_{database}(l)$ where length of each feature vector is L_f, then the Euclidian minimum distance based error between these two vectors is defined as

$$Error = \sum_{l=1}^{L_f} \left| F_{tset}(l) - F^{m,n}_{database}(l) \right|^2 \qquad (2)$$

In the face recognition system, the classifier is designed using sequential adder and multiplier.

III. Hardware Architecture

The architecture of the proposed design is shown using a flow chart in Figure 1. It is divided into four parts: feature extraction, centroid calculation, template database and classifier. All these four sections are controlled by a 3-bit controller.

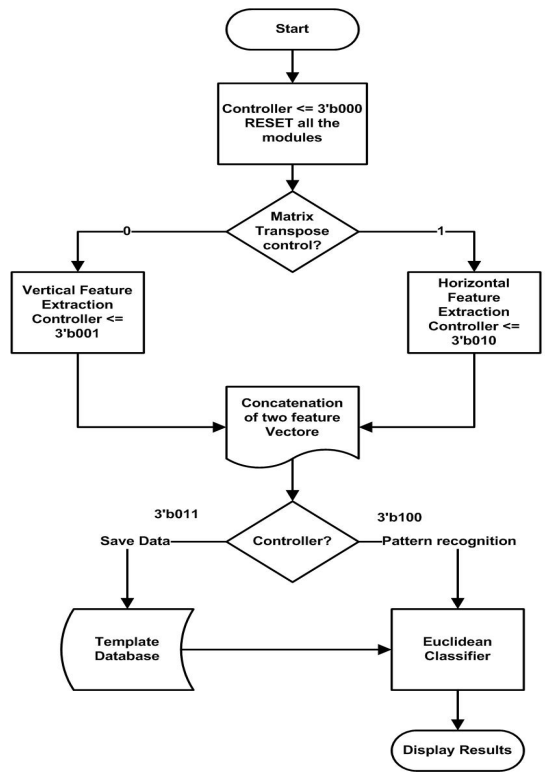


Figure 1: Architecture of the Proposed Face Recognition system

The computation of zero lag cross correlation value is a dot-multiplication of two vectors. Suppose, two vectors are given by:

$$\mathrm{u} = [\mathrm{u}(1), \mathrm{u}(2), \ldots, \mathrm{u}(\mathrm{K})]^{\mathrm{T}}$$

$$\mathrm{v} = [\mathrm{v}(1), \mathrm{v}(2), \ldots, \mathrm{v}(\mathrm{N})]^{\mathrm{T}}$$

where, it is implicitly understood that:

$\mathrm{u(k)} = 0$ for $\mathrm{k} = -\infty, \ldots, -1, 0$ & $\mathrm{k} = \mathrm{K}+1, \mathrm{K}+2, \ldots, +\infty$

$\mathrm{v(n)} = 0$ for $\mathrm{n} = -\infty, \ldots, -1, 0$ & $\mathrm{n} = \mathrm{N}+1, \mathrm{N}+1, \ldots, +\infty$

Now if we want to compute the cross correlation, which can be described mathematically as:

$C(w) = \sum_{n=1}^{N} u(w+n).v^*(n)$ for $w = -N+1, -N+2, \dots, -1, 0, 1, \dots, K-2, K-1$

Where it is known that:

$C(w) = 0$ for $w = -\infty, \dots, -N-1, -N$ & $w = K, K+1, \dots, +\infty$

Where * indicates complex conjugate (of the signal v). This means that there are $(N+K-1)$ cross-correlation values, $C(w)$, to compute. Although implicitly given by zero it should be ensured that all needed u-values are directly given - also the values, which are zero.

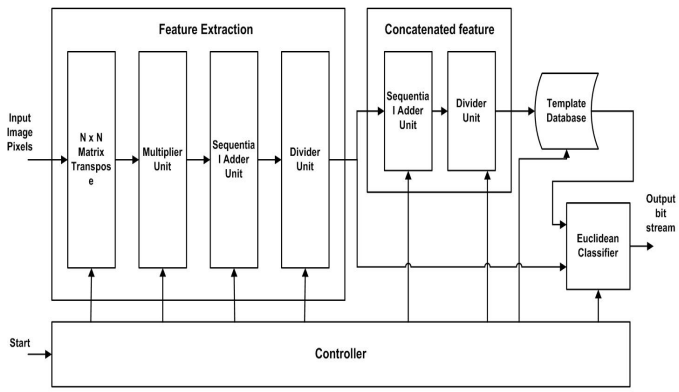


Figure 2: Block diagram of proposed face recognition system

Figure 2 presents the block diagram of proposed hardware. Each module has its own reset and controller. The descriptions of important modules are given in the following section:

A. Matrix Transpose:

This module transposes input matrix into output. Direct pixel values (gray scale image) are the input to this module that are organized as 2×2 matrix. At control zero, output of matrix transpose module is exactly same as the input. In that case, following modules assist the input in such manner that other modules can extract the vertical feature of the input matrix. At control one, it transposes the input matrix and following modules extract the horizontal feature of the transposed matrix.

B. Multiplier:

In the proposed design an 8-bit multiplier is used for multiplying a single 32-bit input bit stream. The output is a concatenation of two resultant bit streams. First the module separates the bit stream into four 8-bit inputs. Then each two 8-bit input multiplies with each other. Then two resultant multiplications are concatenated and send it to the output. The multiplier builds on the concept of positional numeral system. It multiplies the multiplicand by each digit of the multiplier and then adds up all properly shifted results.

C. Sequential adder:

Two sequential adders have been used in the design. One of them performs face extraction algorithm and the other calculates the centroid. The working function of both modules is same. The only difference between these two modules is input bit size. The module that used to design the feature extraction algorithm can handle 16-bit input and the other module can handle only 8-bit input. The structure of sequential adder consists of one 16 bit full adder and a feedback path. At first, it sorts 32-bit bit stream into two 16-bit inputs. The addition of two 16-bit input drives to the feedback path. The feedback path drives the output to the next input for further addition in each clock cycle. It means the output is added with the next upcoming 16-bit input in each clock cycle. In the beginning the input adds up with zero as the modules is in reset mode and saves into register temporarily at the output for feedback purpose. Thus every output of the sequential adder is added with the next upcoming input bit. The results of the sequential output do not go over 24-bit.

D. Divider:

Like sequential adder, two divider modules have been employed - one for feature extraction and the other for centroid calculation. The only difference between these two modules is the divisor number. The first module's divisor is 92 and 112 decimal number depending on which feature vector is going to be extracted. Another divider module's divisor number is 10. It normalizes the output by taking most significant 8-bit number from the quotient. Here the input bit size is 24-bit and output bit size is 8-bit.

E. Classifier:

In this hardware design, a classifier is required for template matching process. Keeping computational burden in mind, simple distance based Euclidean classifier is used in this hardware. Equation (2) represents minimum distance-measuring algorithm. Subtraction is done using 2's complement and multiplication is performed using addition and shift. If the resultant bit stream of classifier is close to zero it means it matches with the template. The opposite of that indicates the differences between training and testing phase.

IV. RESULT & DISCUSSION

In the proposed hardware, Cross-correlation method has been used for feature extraction and minimum distance based Euclidean classifier has been used for template matching process. Extensive simulation has been carried out in order to test the face recognition performance obtained by the proposed method using MATLAB. The entire test has been performed using ORL database [15]. The ORL database has been used extensively in many literatures regarding face recognition. It contains a total of 400 gray scale images, each of size 112×92. There are 40 persons, including both males and females, and each of them has 10 different poses. All the images are taken against a dark homogenous background, little variation of illumination, different facial expression and details. In order to immunize the algorithm to variation of illumination, dc part of the sequence is removed. The low frequency components or the dc part of an image carries the brightness information and high frequency components or the ac part of an image carries the information of sharp contours and crisp edges. We are mostly focus on the changes of information on vertical and horizontal direction. That is why the dc part of the sequence

has been removed. It has been done by subtracting the sequence values from mean of sequences. The Euclidean distance based classifier has been employed in order to test the recognition accuracy. The recognition accuracy has been justified on three different cases, such as concatenated feature vector, only horizontal feature vector and only vertical feature vector. The experiments were performed following the leave-one-out cross validation rule.

Figure 3 represents images of a single person with different facial expression. Figure 4 represents concatenated feature vector of vertical and horizontal feature vector. It also proves the class compactness of vertical and horizontal feature vector. Figure 5 provides a better visualization of the effectiveness of this simple yet effective algorithm. It is clear that the horizontal feature vector is less effective at uniquely representing an individual than the vertical feature vector. Using the combined feature vector leads to an impressive 98.3% accuracy.

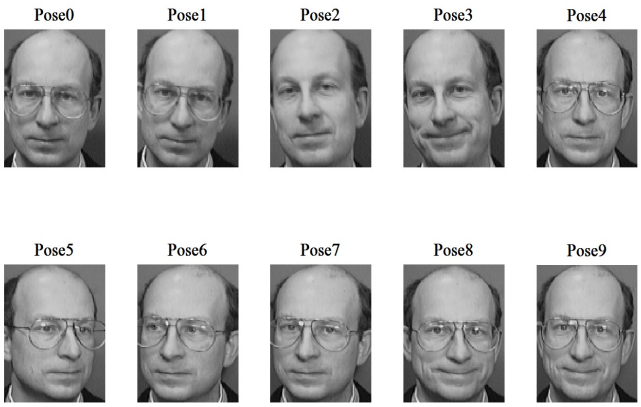


Figure 3: Images of a single person with different facial expression (Collected from ORL database)

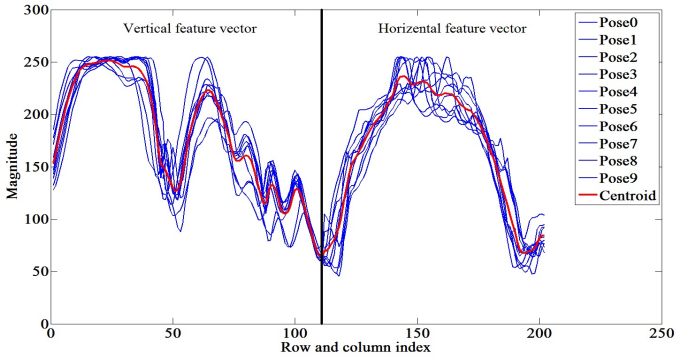


Figure 4: Concatenated feature vector of vertical and horizontal feature

Then the proposed face recognition system is modeled in Verilog and simulated for functional verification in ModelSim. The architecture has been synthesized using Synopsys's Synplify premier for different FPGA device. The synthesized results are shown in Table 1. It should be noted that these results give us a pre-assessment of the resource utilization of the proposed face recognition hardware. All the comparison results are validated with the ORL database with 400 images of 40 subjects and 112×92 size in each gray level image. Input pixels are considered as bit streams and limited to 0 to 255. The maximum clock frequency was found to be 184.7 MHz in Xilinx Virtex7 FPGA.

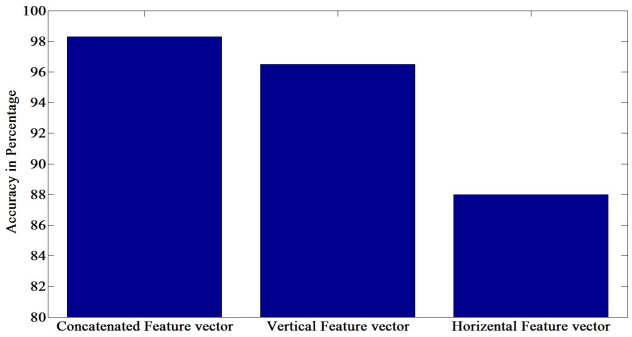


Figure 5: Recognition accuracies of concatenated, vertical and horizontal feature vector

TABLE 1: SYNTHESIS RESULTS FOR DIFFERENT FPGA DEVICES

FPGA device		Area Utilization		Max. Requested Clock Frequency (MHz)
		Register Bits (%Utilization)	*Total LUTs (%Utilization)*	
Xilinx	Virtex2p, XC2VP2	146 (5%)	548 (19%)	105.2
	Virtex4, XC4VLX25	148(0%)	535 (2%)	89.5
	Virtex5, XC5VSX50T	141 (0%)	366 (1%)	149.4
	Virtex7, XC7VX485T	141 (0%)	391 (0%)	184.7
Altera	Cyclone-V, 5CGXFC7C6	135 (0%)	531 (1%)	183.1
	Cyclone-IV, EP4CGX50C	135 (0%)	400 (1%)	97.7
	MAX-II, EPM1270	135(1%)	593 (46%)	128.7
	Stratix-III, EP3SE50	135 (0%)	406(1%)	170.3

Table 2 shows a comparison of the proposed work with similar designs [15]-[18]. For a fair comparison, we have implemented our design in the same technology as the compared works. All comparisons are performed based on commendable feature extraction method and classification. Face accusation has not been considered during the comparison because our proposed hardware has no separate face accusation system. From the observation it is clear that the proposed design performs better on aspect of minimum area utilization and clock frequency. Among all of them, the proposed design has the highest operating frequency. The testing of any image using the proposed system provides a decision with an accuracy of 92.7%.

TABLE 2: HARDWARE IMPLEMENTATION COMPARISONS

Technology	Reference design	Area Utilization		Max. Clock Freq. (MHz)
		Register Bits (%Utilization)	*Total LUTs (%Utilization)*	
Xilinx Spartan-2E	[15]	-	292	50
	Proposed Design	82	263	44.6
Xilinx Spartan-2E	[16]	652	830	50
	Proposed Design	82	263	44.6
Xilinx Spartan-3	[17]	128	256	-
	Proposed Design	87	294	51.8

V. CONCLUSION

In this paper, our objective is to develop a cost effective feature extraction hardware, which can provide high face recognition accuracy. In this regard, we intend to adopt the holistic approach for precisely capturing the variations in the whole image. Zero lag cross-correlation operation is performed considering pairs of consecutive rows and column of an image data to extract the vertical and feature vector. Matrix transpose, multiplier, sequential adder and divider, these are the main modules that have used to design the proposed face recognition hardware. Recognition accuracy of proposed system has been testified. At the same time proposed hardware design has been synthesized using Synopsys's Synplify premier for different FPGA device to justify which platform would be better for implementation based on area utilization and maximum clock frequency. The design best suits for Xilinx Virtex7 XC7VX485T. In addition, synthesis results of proposed hardware design have been compared with some other related synthesized results and the results are satisfactory. The testing of any image using the proposed system provides a decision with an accuracy of 92.7%. All the comparison results are validated with the ORL database with 400 images of 40 subjects and 112x92 size in each gray level image.

REFERENCES

[1]. Z. M. Hafed and M. D. Levine, "Face recognition using the discrete cosine transform," International Journal of Comoputer Vision, 2001, vol. 43, no. 3, pp. 167–188, 2001

[2]. F. M. de S. Matos, L. V. Batista, and J. v. d. Poel, "Face recognition using DCT coefficients selection," ACM symposium on Applied computing, 2008, pp. 1753–1757

[3]. W. Zhao, R. Chellappa, A. Rosenfeld, and P. J. Phillips, "Face recognition: A literature survey," ACM Computing Surveys (CSUR), 2003, vol. 35, no.4 pp. 399–458

[4]. M. Turk and A. Pentland, "Eigenfaces for recognition," Journal of Cognitive Neuroscience, 1991, vol. 3, no. 1, pp. 71-86

[5]. M.S. Bartlett, J.R. Movellan, and T.J. Sejnowski, "Face recognition by Independent Component Analysis," IEEE Transaction On Neural Networks, 2002, vol. 13, no.6, pp. 1450-1464

[6]. Himaanshu Gupta, Amit K Agarwal, Tarun Pruthi, Chandra Shekher, Rama Chellappa, "An Experimental Evaluation of Linear and Kernel-Based Methods for Face Recognition," Application of Computer Vision, 2002, pp. 13-18

[7]. J. Kittler, A. Hilton, M. Hamouz and J. Illingworth"3D Assisted Face Recognition: A Survey of 3D Imaging, Modeling and Recognition Approaches," Computer Vision and Pattren Recognition – Workshops, 2005, pp. 114

[8]. Ara V. Nefian and Monson H. Hayes III, "Face detection and recognition using Hidden Markov Model," International Conference on Image Processing, 1998, vol. 1, pp. 141-145

[9]. Niu Liping, Zheng Yanbin, Li Xin Yuan and Dou Yuqiang, "Bayesian Face Recognition Using 2DPCA," Information Technology and Application, 2009, vol. 2, pp. 567-570

[10]. Yun-feng LI, "A Face Recognition System Using Support Vector Machines and Elastic Graph Matching," Artificial Intelligence and Computational Intelligence, 2009, pp. 3-6

[11]. N. Shams, I. Hosseini, M.S. Sadri and E. Azarnasab, "Low Cost Fpga-Based Highly Accurate Face Recognition System using Combined Wavelets with Subspace Methods," IEEE International Conference on Image Processing, 2006, pp. 2077-2080

[12]. P.N Belhumeur, D.J Kriegman and J.P. Hespanha, "Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection," IEEE Transection on Pattern Analysis and Machine Intelligence, 1997, vol. 19, no. 7, pp. 711-720

[13]. M. S. U. Sarwar, A. Sharin, M. R. Khan, H. Imtiaz, and S. A. Fattah, "A Face Recognition Scheme Based on Spatial Correlation Function," TENCON 2010-2010 IEEE Region 10 conference, 2010, pp. 671-674

[14]. S. A. Fattah, M. R. Khan, A. Sharin and H. Imtiaz, "A face recognition scheme based on spectral domain cross-correlation function," TENCON 2011-2011 IEEE Region 10 Conference, 2011, pp. 10-13

[15]. "The ORL database of faces," 2007 [online], Available: www.cl.cam.ac.uk/Research/DTG/attarchive/pub/data/

[16]. Fab Yang and M. Paindavoine, "Implementation of an RBF Neural Network on Embedded Systems: Real-Time Face Tracking and Identity verification," IEEE Transactions on Neural Networks, 2003, vol.14, no. 5, pp. 1162-1175

[17]. E. A. Abdel-Ghaffar, M. E. Allam, H. Mansour and M. A. Abo-Alsoud, "A Secure Face Recognition System," International Conference on Computer Engineering & Systems, 2008, pp. 95-100

[18]. S. A. Dawwd and B. S. Mahmood, "A reconfigurable interconnected Filter for Face Recognition based on Convolution Neural Network," 4th International Design and Test Workshop (IDT), 2009, pp. 1-6

Password Construction and Management Strategies of the Online Users of Bangladesh: A Demographic Comparison with the Users of the First-World Countries

S M Taiabul Haque
Department of CSE
University of Texas at Arlington
Texas, USA
Email: eresh03@gmail.com

Tauhidul Alam
Department of CSE
Chittagong University of Engineering & Technology
Chittagong, Bangladesh
E-mail: tauhid_cuet@cuet.ac.bd

Mamoon-Al-Rasheed
Department of CSE
State University of Bangladesh
Dhaka, Bangladesh
Email: mamoon@sub.edu.bd

Matthew Wright
Department of CSE
University of Texas at Arlington
Texas, USA
E-mail: mwright@cse.uta.edu

***Abstract*—In this paper, we report on a study of the password construction and management strategies of advanced online users in Bangladesh. The study involved a survey of 75 undergraduate Computer Science students from two different universities. Their responses suggest that even these advanced online users are not aware of some generally accepted good password practices. Our findings are also compared with the responses of various user studies that have been conducted among the users in first-world countries. The Bangladeshi users in our study, who have fewer passwords than first-world users, are more likely to write down passwords and share passwords less frequently. They also do not change their passwords regularly. On the other hand, users in our study generate passwords with fewer personally meaningful words and numbers than first-world users, and a majority of our participants do not use common English or Bengali words. The study is the first of its kind in Bangladesh and it highlights the key weaknesses in the password management strategies of these advanced online users.**

Keywords—Password, usable security, user study, demographic comparison

I. INTRODUCTION

In the present age of computer and information technology, creating and maintaining an online account has become an important task. A user needs to create an online account in a website for the purpose of being exclusively identified by that website. This online account gives the user a unique identity, through which the user can interact with the website authority or other users. An online account acts on behalf of a user in today's digital world and the identity of a user can easily be compromised by compromising any important online account of that user. Therefore, the security and protection of an online account is of paramount importance. In most of the cases, this security relies upon remembering and protecting a sequence of characters, which is also known as password.

Although other forms of authentication mechanism exist, password-based authentication is considered to be the most cost effective one. Several researchers have proposed biometric authentication schemes by exploiting fingerprints [1], retina of eye [2], or voice properties [3]; but their implementations require a lot of expenses. Passwords can be graphical and textual but graphical password schemes have limited deployment in real world. Graphical password schemes like hand-drawn doodle passwords [4] or image-based passwords [5] are not as popular as textual passwords due to the simplicity associated with the use of textual passwords.

Textual passwords, however, are not free from limitations. A good textual password needs to be an easy-to-remember and a hard-to-guess sequence of characters [6], which sometimes presents a dilemma. Because words that are easy to recollect from memory are generally single and short words that can be found in a standard dictionary. If this kind of word is chosen as a password, the password becomes susceptible to dictionary attacks [7], which attempt to guess a password by systematically entering every word in a dictionary as a password.

This password management problem is aggravated when a user needs to maintain multiple password-protected accounts. A recent study revealed that on average, a user of a first-world

country maintains 25 password-protected accounts [8]. On the other hand, in an earlier work, it was reported that the cognitive capacity of an average user restricts the user from dealing with more than four or five passwords effectively [9]. As a result, an average user uses the same password for multiple accounts. This habit is called "password reuse" and if a user reuses a password across multiple accounts, a hacker gaining access to one of the accounts may also be able to compromise other accounts. So, reusing the same password across multiple accounts opens the door for cyber criminals for exploiting a user. This "domino effect of password reuse" has been discussed in detail in [10].

Apart from password reuse, users also follow other poor password practices. Survey results have reported that users write down their passwords on a piece of paper for the convenience of remembering them [11], they frequently share passwords with other people [12], and they use personally meaningful words or numbers that are easily guessable [13]. Due to the lack of consciousness of users, security experts now acknowledge that: "..security is only as good as it's weakest link, and people are the weakest link in the chain." [14]. As a result, numerous user studies have been conducted by modern cyber security researchers to explore the underlying human factors of computer security; and a contemporary branch of security has evolved, which is known as "usable security".

In this study, we conducted a survey among the users of Bangladesh in order to get an overview of their password practices. To the best of our knowledge, this is the first academic study conducted in Bangladesh about password practices of users. We compared the responses of our survey with the responses of the surveys that have been conducted among the users of the first-world countries and highlighted the key weaknesses in the password management strategies of the users of Bangladesh.

The rest of the paper is organized as follows. In Section II, we describe the methodology of our study. We present our findings in Section III. Section IV discusses about the implications of our findings. We summarize our key findings and highlight about future research direction in Section V.

II. Methodology

We conducted an online survey with 31 undergraduate students (22 male, 9 female) of a private university located in Dhaka (the capital and the largest city of Bangladesh) and 44 undergraduate students (39 male, 5 female) of a public university located in Chittagong (the second largest city of Bangladesh).

The survey was conducted through Survey Monkey. Survey Monkey is a website that helps to administer online surveys. We conducted a pilot study among five students first. The pilot study revealed that the students felt more comfortable if the questions were asked in Bengali, their mother language. Therefore, in the actual survey, we asked the questions in Bengali. But we also provided the English translation for each of the questions and for all the answer options.

The survey consisted of ten questions. Questions were asked about the frequency of password-protected accounts, password reuse strategy, password sharing habit, languages used while creating a password, frequency of changing passwords, and password construction and management methods. All of the questions were closed-ended multiple choice questions. We used a 5-point Likert scale (1=Never, 2=Seldom, 3=Sometimes, 4=Often, 5=Always) for some of the questions. Demographic questions were included at the end of the survey. For privacy and security reasons, we only asked the participants about their gender and age. We did not ask for any personally identifiable information like name or email address. In this way, we ensured the privacy and security of the participants of our survey. The students did not receive any direct benefit (money or extra course credit) for participating in the survey. However, we told them that by participating in our survey, they would be able to understand their own password behavior more clearly.

One of the main objectives of our study was to compare between the password practices of the users of Bangladesh and the users of the first-world countries. Therefore, we sought a study population that is as web-savvy as average Internet users of the first-world countries. Therefore, we conducted our study among the Computer Science majors. These users should possess more online accounts than an average user of Bangladesh and they should also be more conscious about the security of their online accounts. By limiting our study sample to Computer Science majors only, we tried to make the comparison as fair as possible.

III. Results

We now present our major findings from the survey. Three of the participants did not complete the survey and we did not consider their responses in our results.

A. *Number of Password-protected Accounts*

Since it is really hard for a participant to count the exact number of his/her password-protected online accounts, we asked the participants to specify the range. The responses of the participants are shown in Table I.

B. *Password Reuse Habit*

When the participants were asked about their password reuse habit, their responses suggested that they frequently reuse a

TABLE I: Number of Password-protected Online Accounts

Number of Accounts	Number of Participants
Less than 5	22
5 to 10	24
11 to 15	11
16 to 20	3
Greater than 20	12

TABLE II: REASONS CITED FOR PASSWORD REUSE

Reason	Number of Participants
I have too many accounts	3
It is easier to remember	37
I reuse the same password for the accounts of same category	6
I only use one password for all of my accounts	3
I reuse the same password for all unimportant accounts	7
I never reuse the same password for multiple accounts	16

password for multiple accounts. Only 22.2% (16 out of 72) of participants reported that they "never" reuse a password for multiple accounts. On the other hand, 8.3% (6) reported that they "seldom", 41.7%(30) reported that they "sometimes" and 19.4% (14) reported that they "often" reuse a password. Surprisingly, 8.3%(6) of participants reported that they "always" reuse a password for multiple accounts.

The participants were also asked to cite the reason for reusing a password. Their responses are summarized in Table II.

C. Password Construction Language

Participants were specifically asked whether they use English words/phrases in their passwords or they construct their passwords by using Bengali words/phrases. We also gave them a third option "While creating my password, I use random words/phrases more that are not associated with a specific language (for example – abd123, asdf etc.)".A majority of the participants (37 out of 72) selected this third option. On the other hand, only 37.5% (27) of participants reported that they use English words/phrases, while only 11.1% (8) of participants reported that they use Bengali words/phrases. This revealed an important implication about the password security of the users of Bangladesh. This issue is discussed in detail in Section IV.

D. Password Sharing Habit

We asked the participants how frequently they share their Facebook or email passwords with friend and family. Almost 80% of participants (70.8% said "never", 8.3% said "seldom") indicated that they do not share passwords. This suggested that password sharing is not a widespread habit among the users of Bangladesh.

E. Password Construction Methods

Since we were interested to know about the password construction methods of the users, we specifically asked the participants what practices they follow while constructing a password for an important account. We gave them a series of six practices and asked them which of the practices they follow. The responses of the participants are summarized in Table III. The first four practices were related with the concept of the bit strength of passwords. The bit strength of a password can be calculated as $\log_2((\text{alphabet size})^{\text{length}})$ [8].

TABLE III: FACTORS CONSIDERED WHILE CREATING A PASSWORD (MULTIPLE RESPONSES ALLOWED)

Practices	Frequency
I use both capital and small letters	32
I use digits	33
I use special characters	23
I create a password longer than 7 characters	36
I use personally meaningful words (school name, birthplace name etc.)	16
I use personally meaningful numbers (age, birthdate, student id etc.)	15

Here, the alphabet size is the sum of the sizes of the different types of characters that are used in a password. These types and sizes are lowercase letters (26), uppercase letters (26), digits (10), and special characters (22). Thus, the bit strength of a password increases if more uppercase letters, digits and special characters are used in that password. The bit strength also increases with the increase of the length of the password. For example, an eight character password that contains both uppercase and lowercase letters and special characters would have bit strength $\log_2(74^8) \approx 49.68$.

The time that is required to crack a password by a password cracker is related to the bit strength of that password. A password that has higher bit strength (a long password with a combination of uppercase and lowercase letters, digits and special characters) is considered to be a good password. It is harder for a password cracker to crack that password because the password cracker would require generating and testing more unique permutations for cracking that password.

It can be seen from Table III that 44.4% of participants of our study reported that they use capital letters and 45.8% of participants reported that they use digits in their important passwords. The frequency of using special characters is comparatively low. About 31.9% of participants reported that they use special characters. However, exactly half of the participants reported that they create a password longer than seven characters for an important account. These findings suggested that for an important account, most of the users create a password that has a reasonably high bit strength.

The last two practices are related with personal experiences and preferences (Table III). The responses of the participants suggested that most of them do not use personally meaningful words (school name, birthplace name etc.) or numbers (age, birthdate, student id etc.) while creating an important password.

F. Frequency of Changing Passwords

We asked the participants how frequently they change their passwords. Table IV shows their responses.

For this question, we allowed the participants to give their own answers. Seven of the participants wrote their own answers. Four of them indicated that they do not have any specific time duration for changing their passwords, while three wrote that they only change their passwords in case of emergencies (spam or virus attacks) or if they feel that their passwords have been stolen.

TABLE IV: FREQUECNY OF CHANGING PASSWORDS

Frequency of Changing Passwords	Responses
Never	24
Weekly	3
Monthly	6
Every three months	7
Every six months	10
Yearly	15

G. Password writing habit

The participants were also asked about the habit of writing down their passwords. Most of the participants revealed that they do not write down their passwords on a paper or in a computer file for the convenience of remembering them. Writing down the passwords on a piece of paper or in a computer file is a bad habit because the paper or the file may get into wrong hands. More than half of the participants (37 out of 72) reported that they "never" write down their passwords on a piece of paper or in a computer file and 20.8% (15) reported that they "seldom" do so. On the other hand, 20.8% (15) and 2.8% (2) reported that they "sometimes" and "often" write down their passwords, respectively. Only three participants reported that they "always" write down their passwords on a paper or in a computer file.

IV. DISCUSSIONS

We explicitly recruited undergraduate Computer Science majors to participate in our survey. Therefore, we do not claim that our results are the best representation of the general Internet users of Bangladesh. However, as mentioned before, we intentionally recruited participants from a technologically savvy demographic because we wanted to explore the password management strategies of the users who are considered to be more educated about technology than an average Internet user of Bangladesh. Our results reflected a best-case example of the users of Bangladesh. It identified the poor security practices that are followed by the users who are relatively more knowledgeable about Internet security and privacy. Our survey results demonstrated that even these users are not well-aware of some of the good password practices. The general Internet users of Bangladesh cannot be expected to be more conscious than this sample population. Our results, therefore, suggest that pragmatic steps should be taken to make the users of Bangladesh more security-concerned about password management strategies.

Our study also highlights several important demographic issues. A recent study from Microsoft Research showed that on average, an Internet user of a first-world country has 25 password-protected accounts [8]. Compared to that, advanced Internet users (Computer Science majors are considered as advanced Internet users because there should be easy Internet access in their labs and they should spend more time on computer/Internet than ordinary users) of Bangladesh have fewer password-protected accounts (Table I).

Although the users of Bangladesh have fewer password-protected accounts and they need to deal with fewer passwords, their password reuse rate is not drastically different from the users of the first world countries. A recently conducted study in Carnegie Mellon University reported that more than 80% of the participants reuse a set of passwords for different accounts [11]. Our survey responses indicated that almost 70% of participants (41.7% said "sometimes", 19.4% said "often", and 8.3% said "always") reuse the same password for multiple accounts.

The users of Bangladesh write down their passwords more frequently than the users of the first world countries. Overall, 51.4% of participants of our study reported that they "never" write down their passwords on paper or electronically, while 87% of participants of the study of Carnegie Mellon University reported that they "do not" write down their passwords [11].

Regarding the habit of password sharing, the users of Bangladesh follow this habit less frequently than the users of the first-world countries. According to a recently conducted study by a researcher of Nokia Research Center, one third of the users share their personal email password; while a quarter share their Facebook password [12]. On the other hand, almost 80% of participants of our study indicated that they do not share their Facebook or email password with friend and family.

Compared to the users of the first-word countries, the users of Bangladesh less frequently use personally meaningful words and numbers while creating their passwords. Table III shows that only 22.2% of participants of our study reported that they use personally meaningful words, while 20.8% of participants reported that they use personally meaningful numbers. In a similar study that was conducted with the students of Wichita State University, a public university of USA, it was indicated that 54.9% of users use personally meaningful words, while 49.8% of users use personally meaningful numbers while creating their passwords [13]. This demonstrates a good practice that is followed by the users of Bangladesh because if personally meaningful words or numbers are used in a password, the password becomes more guessable by others, especially by the close people because they should be more familiar with the words or numbers that are meaningful to a user.

From the responses of the participants, it was also revealed that when creating a password, most of the users avoid English dictionary words and uses random words/ phrases that are not associated with any specific language. As discussed in Section I, an attacker uses a dictionary or wordlist for successfully performing a dictionary attack. Sophisticated password cracking tools like John the Ripper can assist an attacker in this kind of attack [15]. These dictionaries or wordlists are found for English and most other major languages. The website for the Openwall Project [16], one of the biggest repositories for password recovery utilities, contains wordlists for English and 20 other languages. So, a password is relatively safe if English dictionary words are avoided in that password and the responses of our study

showed that most of the users follow this safe practice of avoiding English dictionary words in their passwords. This suggested that the passwords of most of the users of Bangladesh are less susceptible to dictionary attacks.

V. Conclusion and Future Work

We conducted the first academic study about password behavior among the users of Bangladesh and compared our findings with the findings of the studies that have been conducted among the users of the first-world countries. Our study revealed several important findings:

- Even the advanced Internet users of Bangladesh have fewer password-protected accounts than the average Internet users of the first-world countries.
- Although the users of Bangladesh have to deal with fewer passwords, their password reuse rate is not drastically different from the users of the first-world countries.
- Password sharing is not a widespread habit in Bangladesh as it is in the first-world countries.
- The users of Bangladesh follow the poor practice of writing down their passwords more frequently than the users of the first-word countries.
- Most of the users of Bangladesh do not follow the recommended practice of changing passwords regularly.
- For their important accounts, the users of Bangladesh create passwords that are reasonably strong and they less frequently use personally meaningful words or numbers, which make their passwords less guessable.
- Rather than using English/Bengali dictionary words, most of the users of Bangladesh use random words/phrases that are not associated with a specific language, which in turn make them less vulnerable when dictionary attacks are performed by malicious attackers.

In this study, we presented a general overview of the password habits of the users of Bangladesh. In future works, we will focus more on examining the strengths of passwords that are actually created by the users. We will design experiments that would ask the participants to construct new passwords. Analysis of those passwords might help us to better understand the password construction patterns of the users of Bangladesh.

References

[1] A. R. Roddy and J. D. Stosz, "Finger-features statistical analysis and system performance estimates," Proceedings of the IEEE 85(9), (Sep. 1997), 1390-1421.

[2] C. Marino, M. G. Penedo, M. Penas, and M. J. Carreira, "Personal authentication using digital retinal images," *Pattern Anal. Appl.9,* 1(May 2006), 21-33.

[3] H. –J. Kim, "Biometric is it a viable proposition for identity authentication and access control?," *Computers & Security,* 14(3), 1995, 205-214.

[4] J. Goldberg, J. Hagman, and V. Sazawal, "Doodling our way to better authentication," CHI'02 Extended Abstacts on Human Factors on Computing Systems, 2002, 868-869.

[5] R. Dhamija and A. Perrig, "Déjàvu: A user study using images for authentication," Proceedings of the 9th Conference on USENIX Security Symposium, August 2000.

[6] S. Wiedenbeck, J. Waters, J. -C. Birget, A. Brodskiy, and N. Menon," Authentication using graphical passwords: Effects of tolerance and image choice," Proceedings of the 2005 Symposium on Usable Privacy and Security, July 6-8 , 2005, 1-12.

[7] R. Morris and K. Thompson, , "Password security: a case history", *Commun. ACM 22,* 11 (Nov. 1979), 594-597.

[8] D. Florencio and C. Herley, , "A large-scale study of web password habits", 16th International World Wide Web Conference, May 8-12, 2007, 657-666.

[9] [9] A. Adams and M. A. Sasse, "Users are not the enemy", *Commun. ACM 42,* 12 (Dec. 1999), 40-46.

[10] B. Ives, K. R. Walsh, and H. Schneider, , "The domino effect of password reuse", *Commun. ACM 47,* 4 (Apr. 2004), 75-78.

[11] R. Shay, S. Komanduri, P. G. Kelley, P. G. Leon, M. L. Majurek, L. Bauer, N. Christin, and L. F. Cranor, "Encountering stronger password requirements: User attitudes and behaviors", Proceedings of the 2010 Symposium on Usable Privacy and Security, July 14-16, 2010.

[12] J. J. Kaye, , "Self-reported password sharing strategies", 2011 Annual Conference on Human Factors in Computing Systems, May 7-12, 2011, 2619-2622.

[13] S. Riley, "Password security: What users know and what they actually do",*Usability News,* 8(1),2006.

[14] B. Schneier, "Secrets and lies", John Wiley and Sons (2000).

[15] http : //www.openwall.com/john.

[16] http : //www.openwall.com/wordlists.

.

Types of Autistic Children Identification and Implementation in e-Learning Software

Anika Anwar[1],Ishrat Ahmed[1],Samiha Samrose[1],Hasan Shahid Ferdous[1], Md. Mustafizur Rahman[1]
[1]Department of Computer Science and Engineering, Bangladesh University of Engineering and Technology
Email:{ishratahmedmren, samiha_mumu, anika.anwar, webtonmoy, nahid_cse05}@yahoo.com
Syed Ishtiaque Ahmed[2]
[2]Department of Information Science, Cornell University, NY, USA, Email: sa738@cornell.edu

***Abstract*—Autism is a neurological disorder characterized by widespread impairments in social interactions and communications, as well as severely restricted field of interests, delay in response, or repetitive activities. As autistic children are biased towards their own tastes, it is hard to teach them anything beyond their own type. There are a large variety of types of autistic children, and at the same time it is quite a difficult research issue to find out the behavioral pattern and even harder to bring those under a common platform. In this paper we have focused on finding out the types of autistic children and incorporated the available resources with e-learning. Here we have presented a software package - 'Types of Autistic Children Identification and Implementation' that detects fields of interest for the autistic children and use the information of their own behavioral pattern to assist them flourish in their day to day social life.**

I. Introduction

Autism Spectrum Disorder (ASD) is a lifelong behavioral impairment mainly with abnormalities of socialization and communication, specialized field of interests, delay in response, or repetitive activities. An autistic child begins to show signs of autism at an early age, usually by the age of three. All these make it difficult for them to react in a usual way, also ensuring their education is another big factor since they cannot cope up with traditional learning techniques. So in order to help them overcome their limitations and flourish their potential, it is a must to ensure that proper techniques are applied.

E-learning specialists have identified computer games as an easy way of learning things through joy. Autistic children also generally show interest in playing games but only those that match their interest fields. So to teach them in a playful manner, games can be very handy if designed with their personal choice. With a view to enhancing the skills of autistic children, this experiment has focused on some issues related to the nature of autism. Thus the limitations of the autistic children are considered to be ways to improve their abilities. The characteristics of autism that have been focused here are - severely restricted field of interest, attention deficit, problem remembering sequence or pattern, poor vocabulary development, low response to questions, etc.

This paper concentrates on developing the above listed characteristics through interactive computer games especially for those autistic children who have potentials but cannot be taught due to lack of interests. Extensive experiments were done to identify the interest boundaries of autistic children as well as use the information to decrease rest of the limitations stated above. The particular field of interest of an autistic child can be termed as his/her type. Identifying these types are difficult because there are great diversity of types, which is tough to identify efficiently. In this software, we tried to get type information correctly by using pictures each with number of relevant tags. The data collected through the type identification can be used afterwards to make personalized games for autistic children. We have demonstrated three games along with the implementation of 'type identification' software .They are - 1) Puzzle game, 2) Word making game, 3) Banana grams. These games are developed with a view to developing the behavioral pattern of the autistic children as well as applying e-learning technologies for them.

II. Related Work

People with autism have a limited range of imaginative activities. It is agreed that the common individualities of autism are abnormal reaction to input stimuli, lack of human engagement, and inability to generalize between environments [2]. Early stage education might help autistic persons in coping with the theory of mind deficit difficulty in understanding mental states of others. [3]. Computers have been considered as a promising educational aid for autistic persons. One of the first experiments took place in the early 1960s. Colby had some initial success in using computers to instill an interest in speech-related sounds and language in mute children with autism [4]. Nowadays, with the advancement of Information and Communication Technologies (ICT), computer-aided learning or instruction (CAL/CAI) is considered as a key method for handling autism interventions, particularly for young children. According to the literature, autistic persons, children in particular, enjoy interacting with computers since they are experienced as safe [6]. Recent literature reveals that computer systems represent a controlled environment with minimum or no distractions that is crucial in the education process for autistic children [7]. It has been reported that in interactions with computers, children with autism can feel certain level of control over the environment, which cannot happen when interacting with human beings. Teachers working

with young learners who are autistic, report significant shortcomings with a large number of software platforms that is in use in their classrooms. Therefore, although software design remains poorly documented, several experimental studies have shown the usefulness of computers for autism [11].

III. Observation and Design

A. Why Targeting Autism

As the number of children diagnosed with autism is increasing, interest in their behavioral and educational field seems legit. This substantial increase created interest to many of the researchers to have a deep look into the matter. Various kinds of services are necessary depending on the need of children. These may include therapy, occupational therapy, behavioral therapy, etc. There are many behaviors that general children learn automatically, whereas an autistic child might need special attention. For example, an autistic child may have learnt to count backwards on her own, or she may not recognize objects, pattern matching, or alphabets without special teaching. Many guidelines have been offered regarding their educational objectives, including the development of social skills, use of expressive and receptive language, and communication skills.

B. Why Games

Teaching children has always been a huge challenge. It is proven that any educational word games make it possible for the kids to learn while having fun. If these games can target children of a particular problem, it can be a great way to hold their interest and teach them. It has been verified that the autistic children find the visual interface more interesting than traditional teaching. Addressing their different shortcomings with various games can be of help to develop their skills. The important thing to keep in mind that it is imperative to choose games that are simple, fun and interactive as possible. Games with extremely complicated rules will frustrate them and they will lose interest eventually.

C. Photo Tagging in Personalization Process

Personalized methods of teaching and interaction for autistic children have already been proposed and implemented successfully in various cases. But to make things personal choice based, it is assumed that the likings of the particular persons are known beforehand through non-automated methods like-asking the children's parents, observing the children's behavior for a long time by their instructors etc. How to automate this process effectively has long been a difficult issue. Among the successful communication methods for autistic children, the use of images has been a well-established technique. Images are used as signs to describe various mode of interaction to the autistic people. So here we combine these two things- image tagging and personalization. With the help of this combination, tagged images are used as the tool for assessing the personal biased likings of the autistic children.

D. Why Puzzle Games

The main goal of this game is to help the autistic children in remembering sequence or pattern. This game helps them a lot in stacking or arranging their everyday belongings in a better order. Moreover, it gets them acquainted with computers with a more 'hands-on' approach. After frequent visits to 'Autism Welfare Foundation' in Bangladesh we found that this game is actually played in real-world scenario. This is a very common board game that is played by most of the autistic students. So the children in those institutions have shown much interest than they are with other toys. The main feature that makes this game standalone from several other tabletop puzzle games is that it is designed based on the interest of the autistic children.

E. Word Quiz with Personalized Pictographic Method

Autistic children recognize things better if introduced through pictures or images. Tasks like object introduction or vocabulary building is much easier in the case of autistic children if done with the help of images. Here the word-making game tried to mingle the likings of an autistic child with the object to be introduced. For an example, in some real life cases, to introduce 'bicycle', a picture of it along with the spelling are used. But the fact is, for example- if an autistic child is biased toward to the *yellow* color, then the image of the bicycle to be introduced should be yellow to increase the chances of his/her remembering the word. Keeping this point in mind this word-making game is designed. The specialty of the game is that the same game becomes different for different autistic children with different fields of likings.

F. Why Bananagrams

Bananagrams is a very popular game among the children who has just started to learn building words. For autistic children, this concept of the game is slightly customized for their interest. The idea of the game is to build as many words from given random alphabets. This game addressed those who have passed memorizing the letters and can know some common words. If words are taught with graphical visualization like images in this case - this concept helps a lot. Any visual aid with the corresponding word, affect the children to remember the word. This game is very convenient to teach words of various lengths. Words with various length can be built from a given random letters. It helps them with the idea to choose correct alphabets to form a word. It increases concentration as well as some common vocabulary.

IV. Our Work

There are total four combinations in this software package: 1) Type Identification, 2) Word making game, 3) Jigsaw Puzzle, and 4) Bananagrams. After that, there is an *evaluation statistics*.

A. Type Identification

In this part of our software, we detect the type of individual autistic children by the choices they make. Identifying these

types need intense observation and calculation from the large number of choices s/he makes.

Now it gets to the identification of types phase. To keep track, there is a login session and then a number of pictures are shown to that child. For example, let us imagine that at first the following four pictures of dog, ball, watermelon, and lisianthus flower were displayed. 'dog' has the tags animal, cartoon, and brown, 'ball' has been tagged with toy, white, and black, 'watermelon' has tags fruit, red, green, and food, and the last one 'lisianthus flower' has flower, white, jungle, and garden tags. Now A child can do three things: 1) Select 2) Skip, or 3) Quit.

If s/he clicks on any image, it is *Selected* and corresponding tags are saved in his/her profile. If s/he clicks *Skip*, it is considered to be rejected, meaning none of the four images attracts him/her. This process goes on until *Quit* is selected. Every 'quit' is considered to be the session ending. Here let the child log in and firstly chooses watermelon (in Fig. 1a) and then chooses grapes (in Fig. 1b). After that the child quits. Then the possible type result will be 'Fruit' that is shown in Fig. 1c. This way the whole process run once is considered as a session. The values considered to evaluate possible types are updated after each session.

B. Puzzle Game

As an improvisation of the usual puzzle game, the data from *Type Identification* software were used to give them to reorder their own *type-matched* pictures. Dynamic in nature, it selects its puzzle elements according to the type of the child playing. Fig. 2a shows the starting point of the game, where the player has to enter their name. With the corresponding type that is retrieved from the database, puzzle for that particular type is shown in the next stage (Fig. 2b). Here supposing the type is 'fruit' we have a puzzle for an apple. The actual image is shown in the upright corner of the interface. There is a timer to measure the time of successful building image. If they manage to complete Fig. 2c they are redirected to the main page to choose another level if they want.

The game has three levels- 1) Easy, 2) Medium, and 3) Hard. At level-1, it is the 'Easy' mode. There are four broken pieces to put in places as shown in Fig. 2b. In 'Medium' level and 'Hard' level, there are six and nine puzzle pieces to place in proper position. As an autistic child plays different levels, his/her pattern matching power increases. Thus this game helps to memorize the pattern of some common objects that enhances their mental intellectuality.

C. Word Making Game

Word-making Game is an e-learning process where word formation is taught and introduction to different objects is made in a playful manner. Here in this game, the data of 'Type Identification' software were used to teach autistic children about how to spell words correctly. There are three levels: 1) *Easy Level* - teaches monosyllabic words, 2) *Medium Level* - teaches disyllabic words 3) *Hard Level* - teaches tri-syllabic word. Now let an autistic child of 'Fruit' type starts playing the game and chooses level 'Easy'. From 'Type Identification' software, it is known that his/her type is 'Fruit'. As the chosen level mode 'Easy' is for monosyllabic words, we show two images of fruits and their spellings. In Fig. 3a, 'grapes' and 'apple' images and the spellings come as hints for the level. Now the player has to go from 'start' to 'end' point.

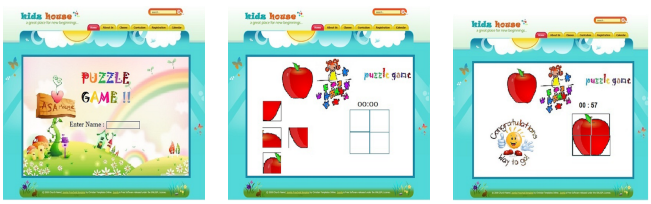
In Fig. 3b the images of grapes and apple are placed on the route points. To pass that route, the spelling is to be written correctly. If he can spell 'grapes' correctly, s/he passes that point (Fig. 3c), and then by spelling 'apple' correctly s/he passes the second route point. There is a 'timer' to evaluate the performance. When an autistic child plays the three levels, s/he gets to know some monosyllabic, disyllabic, tri-syllabic words as well as the corresponding objects.

D. Bananagrams

Bananagrams is a popular game of building words from a set of given random alphabets. Player has to click on the letters to build new words. It is slightly customized for the autistic children so that they are interested to play this game with their own choice. Here in this game, according to the type from 'Type identification' software, autistic children are

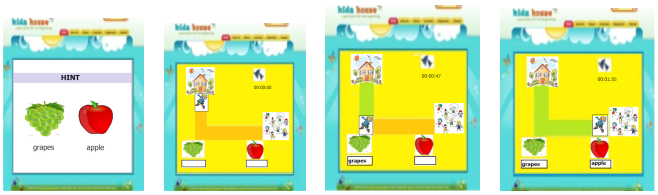
(a) Selection session 1 (b) Selection session 2 (c) Showing results

Fig. 1: Screen shot of type identification software.



(a) Name input Interface (b) Puzzle game GUI at starting point (c) Puzzle game GUI when solved

Fig. 2: Screen shot of the puzzle game.



(a) Hint page (b) Starting page (c) Spelling word (d) Spelling word

Fig. 3: Screen shot of the word-making game.

(a) Starting GUI (b) Name Input (c) Taskpage GUI (d) After successful trial

Fig. 4: Screen shot of the Bananagrams game.

given a set of alphabets to construct words with it. Here, the screenshot shows the user interface for the game Bananagrams. The Fig. 4a navigates to two pages - through the first option, the teacher has the opportunity to enter new words and images to the database. Second one is for the autistic child to start playing the game.

In Fig. 4b player enters name, corresponding type is retrieved from the database and players starts playing the game. Let us say that the type is 'animal'. In Fig. 4c letters for 'monkey' is shown. Now the player has to build word with the alphabets shown in the screen. To make them comfortable with the game, hints are shown above. Then the player has to identify the words that can be generated from those given letters. If they succeed, they are shown the built word with particular image. With this game, they increase the skill to think and re-arrange letters to build different words, which eventually enhances their vocabulary and also to memorize words with visual effects.

V. Experimental Evaluation

The whole package of software has been tested with the autistic children of Autism Welfare Foundation (AWF) and School For Gifted Children and the result is remarkable. The observed result is described below:

A. Type Checking Result

With a view to comparing human observation process with our type identification software to define fields of interest of autistic children, data of both methods are collected. It is seen from Table I that seven out of nine identifications are completely successful, one out of nine results is ambiguous (human observation system also fails here) and one out of our nine identifications is a failure in which case more sessions are to be taken with the software. Table I summarizes the data with provided synonyms of the participants.

B. Puzzle Game Result

To justify the effectiveness of this type specific jigsaw puzzle game, it is divided into three levels - a) Easy, b) Medium, and c) Hard. At a higher level, the number of pieces to reposition increases. After playing a lower level, it is observed that most of the autistic children tend to try a higher level. From the figure, the percentage of autistic children who played 'Easy' mood only is 50 percent. The percentage of children playing both 'Easy' and 'Medium' level is 41.67 percent while percentage of children playing all three levels is 8.33 percent.

This actually proves that those children have become more persistent. Since at different levels performance of the same child can vary, level-based performance can reflect his/her level of persistency at each level. The chart (Fig. 5) shows that overall performance is comparatively well at easy level.

C. Word-making Game Result

While playing this game, often a child required several attempts to finish each hurdle in the game. Each time he/she passed a hurdle, we recorded his/her number of attempts and time per attempt. Finally a performance graph (sessions at a game level vs. time taken per try or attempt) is generated from the saved data. Here in Fig. 6, sessions vs. time at level three taken per try in seconds is shown. The overall result depends on these two points-

- *Session Difference*: It is seen that if the game is played in frequent sessions, then eventually the possibilities of memorizing the word or recognizing the object increase. The more frequent the sessions are, the better the performance becomes.
- *Syllabic Variation*: From the graph the performance comparison among mono-syllabic, disyllabic and tri-syllabic

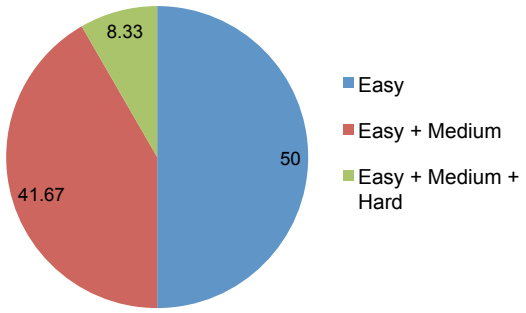


Fig. 5: Level-wise performance in the puzzle game.

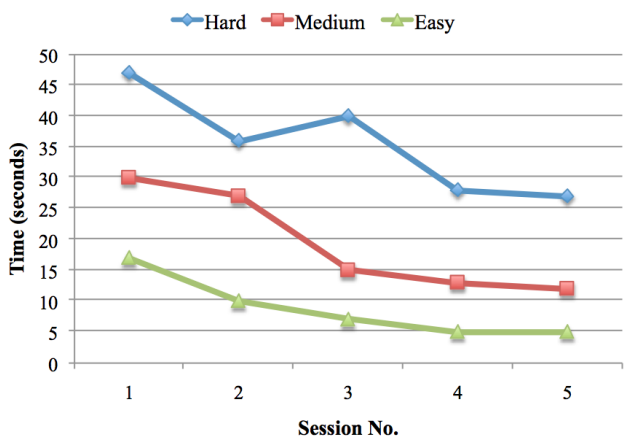


Fig. 6: Time taken per try (in seconds) vs. session number at three levels for word making game.

TABLE I: Comparison between human observed data and data derived from the 'Type Identification' software

Pseudo Name	Type Identified by Human Observer	Type Identified by Software	Note
Roddur	Not Identified, likes random things	Underwater Objects	More observation needed
Shayan	Ball, Balloon, Cars	Toy	Success
Duranta	Decorative Objects	Flower	Success
Anata	Zoo, Pet Animal	Animal	Success
Tahsin	Food	Animal	More sessions needed
Ifty	Anything to eat	Fruit	Success
Poshla	Love to play	Toy	Success
Saad	Animals	Animal	Success
Ayon	Cartoons, Animals	Animal	Success

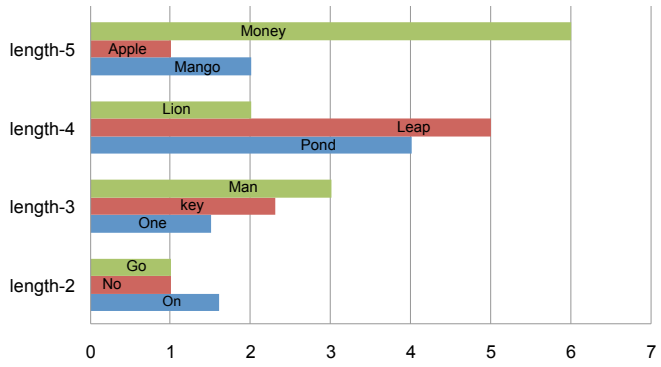


Fig. 7: Horizontal bar chart of word length vs. no of attempts for Bananagrams result.

words is obvious. More to it, as personalized process is applied, the performance of all the levels tends to get better no matter what the levels are.

As a whole, it can be seen that the average number of attempts decreases with increase of the session number. And by maintaining particular choices of the corresponding children, the game can successfully hold the attention of those children in harder levels as well as in easier levels.

D. Bananagrams Result

The main idea of this game is to enable autistic children's power of making words from a set of random letters of a given word. It helps them to think thoroughly as they have to build new words and thus increase their intellectuality. The success of this game was measured by calculating the length of the words that they built in every session. When played by a number of autistic children, two things came out as result :

- They are more comfortable with shorter length of words. They can remember them easily than the longer words.
- If the word is of longer length but that word is known to the player they can keep it in their memory. (For example 'orange' is a word of 6 letters. But as they are familiar with the word, they can easily spell it).

VI. Conclusion

In future, we will add sound clips in the games as an encouragement to play the games. It will also help the autistic children audibility. Add to it, the games will be developed in mobile platform so that they can become widespread and thus help more and more autistic children. The overall target of the software is to improve attention level and incorporate e-learning using personalized information. Moreover playing each the games, an autistic child can develop his/her persistency, vocabulary and responding power. Guardians of an autistic child can plan the personal development process by observing the evaluated result.

References

[1] M. Davis, K. Dautenhahn, C. Nehaniv, S. Powell, *Towards an Interactive System Eliciting Narrative Comprehension in Children with Autism: A Longitudinal Study*. Designing Accessible Technology, 2006.

[2] L. Y. Tsai, *Diagnostic Issues in High-functioning Autism*. in High-functioning Individuals with Autism. New York: E. Shopler and G. B. Mesibov Ed., Premium Press, 1992.

[3] P. Howlin, S. Baron-Cohen, J. Hadwin, *Teaching Children with Autism to Mind-Read: A Practical Guide for Teachers and Parents*. John Wiley and Sons, New York.

[4] K. M. Colby, *The Rationale for Computer-based Treatment of Language Difficulties in Nonspeaking Autistic Children*, Journal of Autism and Childhood Schizophrenia, 1973.

[5] G. Rajendran, P. Mitchell, *Computer Mediated Interaction in Asperges Syndrome: the Bubble Dialogue Program*, Computers and Education, 2000.

[6] D. K. C. Murray, *Autism and Information Technology: Therapy with Computers*, in: S. Powell and R. Jordan Eds. Autism and Learning: A Guide to Good Practice, 1997.

[7] S. J. Green, *Computer-Based Simulations in the Education and Assessment of Autistic Children, Rethinking the Roles of Technology in Education*, Tenth International Conference on Technology and Education, Massachusetts Institute of Technology, Cambridge, MA.

[8] L. Vera, G. Herrera , E. Vived, *Virtual Reality School for Children with Learning Difficulties*. Proceedings of the 2005 ACM SIGCHI International Conference on Advances in Computer Entertainment Technology, 2005, Vol. 265.

[9] A. Luneski, E. Konstantinidis, M. H. Antoniadou, P. D. Bamidis, *Affective Computer-Aided Learning for Autistic Children*. 1st Workshop of Child, Computer and Interaction (WOCCI 08), Chania, Greece, 2008.

[10] M. Barry, I. Pitt, *Interaction Design: a Multidimensional Approach for Learners with Autism*. Proceedings of the Conference on Interaction Design and Children, 2006.

[11] O. Grynszpan, J. C. Martin, J. Nadel, *Human Computer Interfaces for Autism: Assessing the Influence of Task Assignment and Output Modalities*, Conference on Human Factors in Computing Systems (CHI), 2005.

[12] M. V. Panyan, *Computer Technology for Autistic Students*, Journal of Autism and Developmental Disorders. 1984.

[13] D. J. Moore, P. McGrath and J. Thorpe, *Computer Aided Learning for People with Autism - a Framework for Research and Development*, Innovations in Education and Training International, 2000, Vol. 37.

Beetles: A Mobile Application to Detect Crop Disease for Farmers in Rural Area

Rahat Yasir, Nova Ahmed
EECS Department, North South University
Dhaka, Bangladesh
rahat.anindo@live.com; nova.ahmed.2011@gmail.com

Abstract— We have presented our work Beetle designed to support farmers in rural area to detect crop disease. The system shows promises in terms of disease detection. We present our technique here. We have started our software development process with user's requirement analysis. We wanted to complete this cycle by taking Beetles back among the farmers and find out about their opinion on such tool. The farmers received the tool positively and provided us with valuable insights presented in the paper.

Keywords— Plant disease; Mobile Phone Application; Image processing; Human Factors; Design; Measurement; Crop Disease and Fertilizer.

I. INTRODUCTION

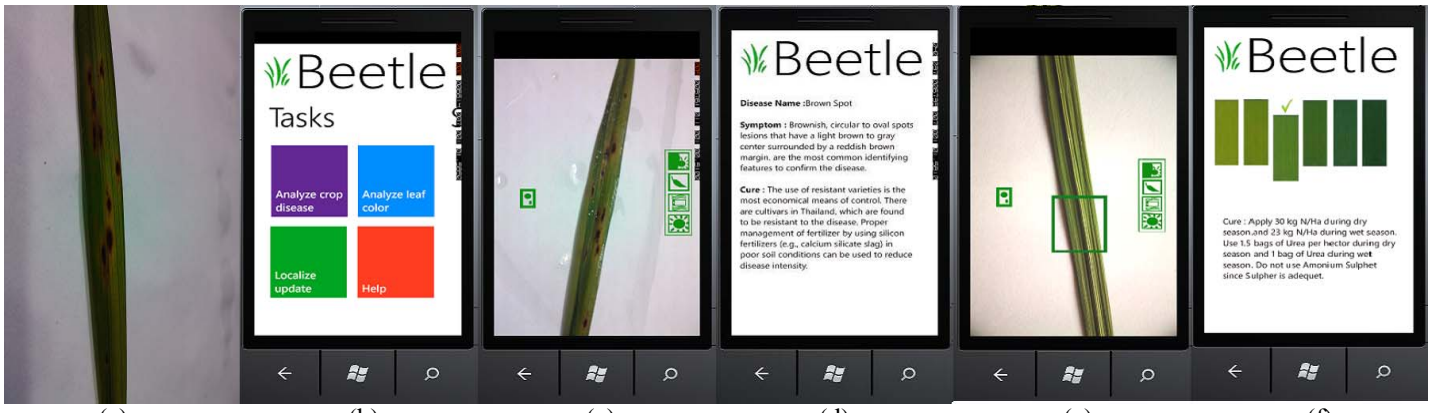
Recent proliferation of cell phone has created unique opportunities in developing countries like Bangladesh. On one side the technology has created communication and compute capabilities all around the country. On the other hand, there is a demand for support at rural places where this cell phones are available. We have focused on an application developed for farmers to detect crop disease which is a major problem among farmers. We developed a cell phone based application "Beetle" that detects crop disease from the image captured by a cell phone and detects the disease in real time. We have studied Beetle among a small group of farmers in our evaluation process along with the evaluation of the software performance itself. Disease detection is a major concern among farmers in the rural areas of Bangladesh. They make their decision based on experiences and assumptions. We have focused on image processing based techniques to detect crop disease. Current image processing algorithms require high quality images [1, 2, 5, 6] which is not the case for cell phone based images. The acquired image must be processed within the cell phone to serve farmers in absence of internet connectivity. We have developed an application that is able to detect crop disease within the limited compute capabilities of cell phone using histogram and color information of captured image. Similar approach has been proposed by [6] to detect the shapes of medicinal plants unlike our approach. There has been various cell phone based applications [add] focused to support rural people. Our approach is complementary to other cell phone based applications as our approach can gain advantage of internet connectivity in terms of information sharing. However, Beetle is able to perform independently. Beetle can detect five different crop diseases successfully and they are Narrow brown leaf spot, Bacterial blight, Brows spot, Ufra and Rice Blast. We have conducted user studies before and after our software development process. We wanted to find out about the burning problems for farmers and then we have showed our complete application to the farmers and got their feedback. We focused mainly on the accuracy and performance of Beetle which was well received among the farmers. We got suggestions about improvement of the current application in terms of its interface and features. Beetle shows promises and gives us future direction for a robust application. Our user study provided us with insights about desirable features such as use of native language and familiarity. The rest of the paper is organized as follows, we present our system design in next section, followed by the results, related work and finally, the conclusion.

II. SYSTEM DESIGN

We present our system design in two phases. We first present the user interface which is carefully designed to be simple keeping in mind the low literacy (or lack of literacy) level among farmers followed by the actual system development process.

A. User Interface

We have focused on a simple interface so that the users can select their options in minimal number of steps. Beetle has two major parts – one is disease detection and then it suggests for in suggestion for fertilizers. The disease detection process goes through four steps. First, we need to place the leaf of observation on against a white background and take a photo using a single click in the current smart phone. Then we may choose our desired step from the list of tasks named Analyze Crop Disease. Then our algorithm uses its internal mechanism to calculate the disease in real time and finally, displays the result in an output screen. Text to speech option is also available for illiterate farmers who are unable to read the output result of Beetle. The fertilizer selection process goes through the first two steps as before. The third step requires the user to select a portion from the leaf as shown in the figure. The fertilizer suggestion is based on the leaf color as shown in Figure 1.



(a) (b) (c) (d) (e) (f)

Figure 1: User Interface for Beetle (a) Leaf Placement (b) User Options of Operations (c) Photo Capture for Disease Detection (d) Output for Disease Detection (e)Photo Capture for Fertilizer Selection (f) Output for Choices of Fertilizer.

B. User Interface

The system considers two major algorithms. The YCbCr algorithm and the Histogram algorithm. We discuss the algorithms briefly.

Figure 2: Algorithm for Fertilizer Determination

In this fertilizer suggestion option beetle is taking the snap and then it will convert the RGB image into YCbCr image to find the color code of the crop leaf. This color space separates the luminance (light intensity) and chrominance in the image (differences of two color components) as shown in Figure 3.

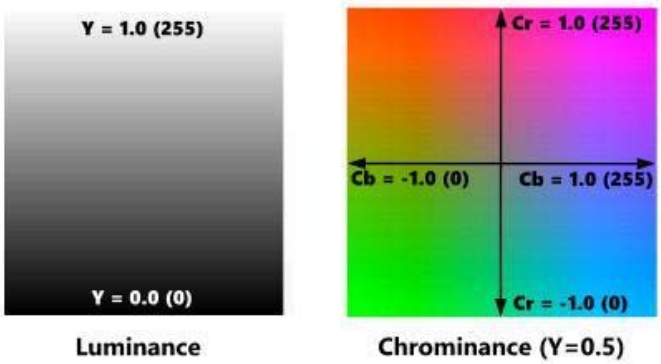


Figure 3: Color Separatin in YCbCr image

Advanced image coding codec converts the source image from RGB to YCbCr color space. The luminance of the image will be stored in the Y channel and Y describes the intensity of the pixels. The Cb and Cr channels hold the chrominance values in blue and red directions. All three have the same amount of contrasts and information. Y channel is perfect but Cb and Cr channel are much less in contrast. When we take a photo and acquire a frame from the camera preview buffer that image will become a one dimensional array of byte values. To optimize the performance of the camera, each Windows Phone device "quarter-samples" the chrominance components of the frame. This means that for every 4 bytes of Y information, onloy one byte of Cb and one byte of Cr is used to describe the corresponding 4 pixels of the image. Furthermore, some of the bytes in the array may be unused to allow room for padding (potentially varying between device models). The image is first converted to one dimensional array of byte values. The byte values and pixel information are compared to define a particular color code according to leaf color. This color code determines the amount of fertilizer by comparing it with predefined value in our system as can be seen in Figure 2.

Figure 4: Generating the Histogram Image

The image is scanned in a single pass and a running count of the number of pixels found at each intensity value is kept. This histogram is a graph showing the number of pixels in an image at each different intensity value found in that image. We can find Histogram (intensity) if we take the total number of pixels divided by total number of pixels at the intensity. We generate the probability of every intensity from the value of the histogram and number of pixels which ranges from 0 to 255 in current scenario [$0 \leq k \leq 255$ is CDF(k) = P(intensity$\leq$k)]. We use this intensity information as the coordinate value of 255 points. We divide this coordinates into 12 different phases. Now we divide this 255 coordinates into 12 different phases and add all the coordinates of each phase and obtain 10 different straight lines. We find the value of each angle using its tangent value. We obtain 10 different values which are stored in an array. We consider the first 6 angel or value of that array. We will consider these values as the unique id for each crop disease and will compare it with the stored unique id. At the same time we run YCbCr algorithm here to find out the exact color code of the leaf. If the matching rate of those id is above 80% and the color code of the infected crop matched with the stored color codes then we can tell that we have successfully detect that crop disease through this process and reply the disease name and cure to the user.

III. RESULT

We evaluate our system in the perspective of system performance, detection rate and then in the light of user experience.

Disease Name	YCbCr (second)	Histogram (second)	Result	Image	Detection Rate (%)
Narrow Brown Leaf Spot	2.05s	1.40s	Narrow Brows Leaf Spot		93%
Brown Spot	2.66s	1.32s	Brown Spot		89%
Bacterial Blight	2.34s	1.56s	Bacterial Blight		86%
Healthy Leaf	1.30s	1.15s	No Disease Detected		98%

Table 1: Performance of Disease Detection Algorithm

Our Beetle can detect five different crop diseases and they are Narrow Brows leaf spot, Bacterial blight, Brown spot, Ufra and Rice blast. We have collected 120 infected crop leaf and we have used 60% of them for training purpose, 20% of them for validation and 20% of them for testing purpose. Overall disease detection rate of Beetle is more than 85%.

Crop Name	YCbCr (second)	Total Time	Image	Result
Healthy Rice Leaf	0.72s	0.72s		Leaf Color Chart Range: 3 Fertilizer: 30 kg N/Ha during dry and 23 kg N/Ha during Wet Season

Table 2: Performance of Fertilizer Determination Algorithm

IV. SYSTEM PERFORMANCE

The runtime of the system varies where we have studied the system using the crop disease for rice considering six different diseases. The average runtime along with the worst case averages are considered. For run time, the fertilizer suggestion takes 0.72 seconds and disease detection algorithm takes 1.32 seconds while the worst case situation takes 4.23 seconds (If both algorithm works together).The crop leaf run time algorithm takes less than a second (0.72 seconds) in average and worst cases. The memory consumption of the system is 5.05 MB (.xap file) where the usage of the system during image processing is 35% (20%-22% usage increase. Running time performance of some key algorithms are shown in Table 1.

V. USER EXPERIENCE

We have studied ten users, all of them are full time farmers ranging the age of 35 to 45 and all of them are male participants which is the case of traditional farming in Bangladesh. All the participants had a cell phone of their own which was our motivation behind a cell phone centric application. Among the participants 90% of them believed plant disease to be a serious problem and all of them mentioned a loss which varied from as small as 5% to as high as 20% each year. The average loss was around 9.8% on average. We have taken our mobile phone based solution and let them take a look at how it works to get their feedback. We had a fixed set of questionnaire ranging from Not Useful, Normal, Good and Very good that was noted down by interviewer. We also got verbal comments along with this fixed answer. One of the farmer thought the solution not to be useful and was skeptical about the ability to detect crop disease telling us “I don’t think this software can successfully detect crop disease.” There was a response where a user was indifferent to the solution and mentioned about the cost of a smart phone as “this phone is very expensive “. Among 40% of the users thought the software to be useful (Good in our data collection set) and mentioned practical problems about the cost of the cell phone, difficulty to operate a high end cell phone as main concerns. We got 40% of the users to like the software (Very Good) and they mentioned the cost, other interfaces available in Bengali.

They suggested “If this software suggest us seed and other fertilizer to according to the plant condition and

environmental condition then it will be the complete solution for farmers" and "we need proper training to use such cell phone" as suggestions for improvement. The study with our users gave us a motivation towards improvements that would be impactful to users such as providing possibilities so that multiple users can share a cell phone to support them in information gathering, exchange and proper detection of disease. We also found it useful to have an interface in native language and making the touch screen based interface available to users to get comfortable with.

VI. RELATED WORK

We have looked at work that relates to our work for image processing as well as an easy to use interface for rural people. We have developed this Application of smart phone technology for detecting paddy crop diseases and suggesting fertilizer on windows phone platform. Smart phones contain very sophisticated built in camera which has less focus and zoom capability than our digital camera. In this case the picture quality is not that much high so that edge detection algorithms most of the time fails to find out differences between infected crop spot and branches of leaves. Even smart phones have limited computing capacity so that they could not process a high quality image or run a high processing algorithm. For that reason we have written codes lightly and efficiently so that device could perform its task in less computing process. There are existing work of advanced image processing by Alom. et. al. using Gaussian Mean Based Paddy Disease Segmentation [1], Wang et. al [2] who have used edge extraction, Revathi et. al. [5] Homogenous Segmentation based Edge Detection methods. The methods work great for edge detection. However, we had to take care of the poor image quality where the lighting may not be adequate , uneven disease structure and low compute capabilities of cell phones. Kumar et. al [3] use histogram to detect different tree leaves which is similar to our histogram based idea, however, we have used this technique to detect the leaf disease instead of the leaf structure. Hannuna et. al. [4] uses manual process to determine the leaf color whereas we have used YCbCr algorithm to find the leaf color through an automatic process after taking the image.

There have been work that focus on mobile application based approaches. Amos et. al [8] have developed a system through which farmers, field officers, experts, buyer, doctor, government ministry, banks are connected with each other, could communicate with each other and work together using their internet based mobile application. But internet connection is not available in most of the rural areas of Bangladesh. Our solution Beetle is totally an independent solution. It is not required to have internet connection to use this mobile phone application.

It is evident that there is a burning requirement for innovative applications even for the rural user's especially rural farmers.

VII. CONCLUSION

We have developed a cell phone based application to support farmers residing in rural areas of Bangladesh to detect crop disease. The application is able to detect ten major crop disease in Bangladesh using the cell phone's camera image of the infected crop. Image processing in a resource constrained device like a cell phone is only useful when the processing time is short. Our application is able to detect the disease within five seconds of time even in the worst case scenario. Our user based study provides us with valuable insights on the wide spread possibility of such an application that would improve the lives of farmers. We look forward to extend our application to play a positive role on plant disease detection as well as in the lives of rural residents like farmers.

REFERENCES

1. Md. Zahangir Alom, Hyo Jong Lee, Gaussian Mean Based Paddy Disease Segmentation. 2010 *the 3rd International Conference on Machine Vision* (2010)
2. Beilei Wang, Ying Cao, Huiming Xiao, Huiyan Jiang, Hongjuan Liu, The edge extraction of agricultural crop leaf. PIAGENG 2009: Image Processing and Photonics for Agricultural Engineerin (2009)]
3. Sandeep Kumar.E, Leaf Ccolor, Area and edge features based approach for identification of Indian Medicinal plants. Indian Journal of Computer Science and Engineering (IJCSE), Vol. 3 No.3 Jun-Jul (2012)
4. Sion Hannuna, N. Anantrasirichai, Swarna Subramanian, Suma Prashant, Ashok Jhunjhunwala and C. Nishan Canagarajah, Agriculture disease mitigation system, ICTACT Journal on Communication Technology, Special Issue on Next Generation Wireless Network and Applications, June 2011, Volume 2, Issue2 (2011)
5. Piyush Chaudhary1, Anand K. Chaudhari2, Dr. A. N. Cheeran3 and Sharda Godara4, Color Transform Based Approach for Disease Spot Detection on Plant Leaf. International Journal of Computer Science and Telecommunications, Volume 3, Issue 6, June (2012)
6. P. Revathi, M. Hemalatha, Homogenous Segmentation based Edge Detection Techniques for Proficient Identification of the Cotton Leaf Spot Diseases. International Journal of Computer Applications (0975 – 888) Volume 47– No.2, June 2012 (2012)
7. A.Senthil Rajan, Image Processing Techniques for Diagnosing Paddy Disease. Proceedings of the World Congress on Engineering 2012 Vol II WCE 2012, July 4 - 6, 2012, London, U.K.(2012)
8. Amos Gichamba, Ismail A. Lukandu, A Model for designing M-Agriculture Applications for Dairy Farming, The African Journal of Information Systems, Volume 4 | Issue 4, 10-8-2012

MassFeed for Developing Countries: Our Opinion, Our Voice

Saif Ahmed, Md. Azharul Islam and Hasan Shahid Ferdous
Department of Computer Science and Engineering (CSE)
Bangladesh University of Engineering and Technology (BUET)
Dhaka-1000, Bangladesh
saif.ahmed@csebuet.org, azharul.parvez@gmail.com, hasan.shahid@csebuet.org

***Abstract*— Participation of mass people in a decision making process is undoubtedly a major concern. It has been a prominent research topic to devise platforms for centralizing public opinion in recent years. Although many internet based and mobile based e-services have been developed, they are out of reach of crowd in a developing country like Bangladesh. Because of illiteracy, lack of internet availability, it often becomes impossible to get feedback from masses. Considering the necessity of designing a feedback system, this paper introduces a novel approach: *MassFeed,* to ensure participation of mass people in a convenient and interactive way.**

Keywords- feedback system; mass people; decision making; opinion;

I. INTRODUCTION

In our present world making a decision might impact a lot. There is no substitute of it when the situation arises to do so. The process is appreciated when it is developed based on opinions or ideas from related people. In a word it implies the importance of taking feedback where it is a process in which information about the past or the present influences the same phenomenon in the present or future. Moreover, it is an important tool to collect and re-act people's opinions. Accumulating all of feedbacks and reaching on a decision that reflects the people's thoughts is not an easy task. The more opinions get involved in the feedback process, the more it increases the probability of being mature and acceptable decision. All these situations clearly show the importance of learning the public opinion at a mass level and raise the question of ignoring mass voice.

A feedback system is a very important decision making parameter for any kind of organization. It is the most powerful, useful, cheapest and widely used process for an organization to learn the public opinion. Authority always focuses on more participation in their feedback process to learn the view of people. Though the present systems cover most of the people through their highly interactive and friendly interfacing on web or mobile based feedback taking applications in developed countries, the illiterate and poor people are always being deprived in the developing and under developed regions as they don't afford to use internet. And that scenario basically represents the South-East Asia Region (SEAR) where almost all countries contain more than 80% of total population according to UN survey 2012 [4], don't get opportunity to use internet. We consider 'Bangladesh' as representative country from the SEAR to provide necessary data throughout the paper, without losing any generality. Bangladesh is ranked 3rd among 8 countries in this region in the e-Government ranking [4]. Despite the efforts of Government to provide e-services to its citizens, the internet access is available to only 5% of total citizens and many of those who have access to internet use it seldom [4]. Besides, her literacy rate is only 56.8% where the criteria are to read and write at the age of 15 or over [3]. So, the need for a complete feedback taking system to cover the most of the people and recognize their voices is undoubtedly high. Keeping in mind all those limitations and feeling the necessity to make the feedback system more convenient we propose MassFeed: a system which is capable of making feedback process more interactive and easier. The implications of the proposed system are quite clear. MassFeed provides a platform to the crowd to enlist their voices in the form of information or data which can be further analyzed by data mining methodologies to build up a decision. As we focus on those who don't or can't take part in conventional feedback system, we think to capture their voices by presenting our feedback system with an ease.

The rest of this paper is organized as follows. In Section II, we present related work of existing feedback taking systems. In Section III, we emphasize our motivation behind planning such a different way. MassFeed is explored and explained in Section IV. Test results and future works are discussed in Section V. Finally, Section VI concludes the work of this paper.

II. RELATED WORKS

Many developing countries have now realized the need of a quality system as well as shown their efforts towards covering masses' voice. To capture crowd source opinions, countries like Bangladesh, India and Pakistan have approached by computerized or web based prior solutions with a variety. These include SMS based feedback system [9], mobile voting [6], online feedback system, feedback through social networking sites [7], using GSM mobile for real-time audience feedback [11], SMS based AVR [8] systems for crowdsourcing opinions. Users need to be connected to internet or have mobile devices to broadcast their opinions. Sometimes, they must be registered as a valid user on the system where privacy cannot be obtained. The main shortcomings of web based approach are: unavailability of location based feedback taking service according to the user demand and lack of internet access for majority of people. In India, a SMS based system has been proposed, incorporating with AVR [1] and GSM [2] technology. This system uses a centralized server and GSM [2]

network to store users' feedback and questionnaire about the topic and communicate with the users. The system asks user to cast his/her choice among the available options. People who are interested to participate need to communicate with the system by sending a text message. In this regard user must have cell phone and people with cell phone might not even show interest because of SMS cost. Besides, in case of some audio recorded mobile voting systems lack of covering people as the access channel is only cell phone and voting through mobile phone again has the problem regarding call cost which is an important pruning factor for SEAR region.

One of the important studies in crowdsourcing developing region is done by A. Gupta [10], who identified that the lack of mass level participation due to lack access to computers and internet of poor communities. They presented mClerk [10] which enables the image protocol to send small images via ordinary sms to increase the participation of mass people. Also, they provided the limited adaptability for their proposal even for a low-end mobile phone which had been deployed over a semi-urban region in India.

From the above discussed works we have identified the limitations of the existing systems as it provide limited attention to accessibility and user friendliness which results the illiterate and poor communities to be deprived crowdsourcing. None of these current initiatives for developing region fully investigate the sole reason behind the lack of mass participation and provide them a suitable solution. To take part in any kind of web or mobile based feedback system one must have internet access or own cell phone and be well aware of information technology. In a third world country like Bangladesh where internet has not reached deep into the mass level and with a lot of people who cannot afford to have a cell phone. As a result, we may get confused if we want to know the opinion of people here using existing feedback system, as huge number of people remains uncovered. To hear those voices and understanding the current scenario, in this paper, we propose a system which is not only accessible to all and keeps anonymity of users' feedbacks but also provides the capability of setting questionnaire according to users' demand.

III. MOTIVATION

There are so many smart methods for taking feedback [5] from people, but it turns out to be different for the masses SEAR which is stricken with problems like illiteracy, poverty and so on. Those people cover a huge percentage of the entire population and also are not able to collaborate with the existing feedback systems as they have their limitations with respect to those solutions. Ultimately determining the true result for any topic which was set for taking feedback is going to be difficult and erroneous as the conventional system produces result ignoring a large population.

Normally in developing countries a lions-share of the population is illiterate and under privileged. As a matter of fact many of them are often ignored in the feedback process. Some people are considered educated but they don't have enough technical knowledge. They also make the side heavier who can't take part in the feedback process. Besides, some are capable of participating but still don't as the existing methods lack interest. To proceed on furthermore the comparison of the above mentioned factors are needed to clarify about the problem.

There has been dispute regarding the total number of internet users in Bangladesh. But despite the fact it is noticeable that very few of total population around 5% use internet [4]. Besides the literacy rate is just more than half of total population where anybody just can read and write considered as literate in this calculation [3]. If we compare this with other countries in the world taking a view from (Table 1) we can see that developing regions are way behind than developed arena. If we consider both internet usage and illiteracy combined, we can see Bangladesh on top of the victimized list.

TABLE I. INTERNET USAGE AND LITERCY RATE ANALYSIS (ALL FIGURES ARE IN MILLIONS OR PERCENTAGE)

Country	Population	Internet user	Internet user (%)	Literacy rate (%)
Australia	22.8	17.9	78.3	99
USA	314.8	243.8	77.4	99
UK	62.3	51.2	82.2	99
Nepal	26.6	2.7	10.3	60.3
India	1210.2	125.0	10.3	83.04
Pakistan	181.3	15.9	8.8	54.9
Srilanka	20.3	3.2	15.6	91.2
Bangladesh	152.5	7.5	4.9	56.8

In order to capture the mass opinions in country like Bangladesh we need to re-think to use about the existing solutions. As it is said earlier, our proposed system understands this scenario and intends to hear the mass voices, by presenting a more convenient system with an ease of use. Because of providing a platform for those uncovered people MassFeed handles the shortcomings of current solutions as we can see it from the comparison between existing systems and MassFeed from (Table 2). It shows us the possible categories of users and the capability of capturing them against these categories by web or mobile based solution on the contrary with MassFeed.

TABLE II. COMPARISON BETWEEN EXISTING SYSTEM AND OUR PROPOSED SYSTEM

Population type	Existing Systems (Web/Mobile)	MassFeed
Internet User	Yes	Yes
Cell phone User	Yes	Yes
Literate	Yes	Yes
Illiterate	No	Yes
People with no internet or cell-phone facility	No	Yes
Educated but unaware of technology	No	Yes

In the following section IV our proposed system will be explored and explained.

IV. Feedback System for Mass People

We think covering those people who are normally ignored by existing systems is important for the authority to take the right decision. Our device is designed by the concept of pre-recorded voice. As many of the audience are illiterate or technologically lag behind to cope with the web or mobile based components, we think of asking user to listen. After getting what the question topic or survey is, user also be instructed by our system, as we keep the further instruction to complete the operation on that pre-record audio. As user may need to hear the question or instruction again, we provide the opportunity to play the audio again. Thus our system works and provides the updated feedback statistics at last.

As GSM [2] or SMS based web solution can't be the speeded to the crowd, we use AVR [1] to reach as much as possible towards the end user. To make our device user friendly we employ the pre-recorded audio and proper guidance on the native language. There is a possibility of having such user who prefers to read out the text rather than listening as it takes a bit long time sometimes. Keeping the thing on mind we show the full audio as plain text by Liquid Crystal Display (LCD), which is depicted in the block diagram of MassFeed in figure 1.

Besides, there is another thought that, not every question is required to have two to four options. It can be more than that. In case of affirmative or negative question like "Have you slept last night?", the choice must be within "YES" or "NO", but for more generic and descriptive case like "What is your monthly expense?" , the option varies as different class of people have different figure of that. The more we increase the options the more we can reach the answer where user may fit him/herself. In our device we also handle this situation of providing to set such questions having maximum eight choices. And every option is marked numerically from (1-8) represented by a single push button. There are also two speakers to make heard the question to the user loudly as the device can be placed in a noisy or crowd place.

Our proposed system is basically a hardware implementation so that masses can make their participation noticeable. The device has been developed using microcontroller supported by the followings:

- microcontroller (Atmega32)
- IC 7805 (Voltage Regulator IC)
- 8-Bit Digital-to-Analog Converters (DAC 0800)
- Linear Regulator (LDO) - Single Channel LDO - LM2937
- LM386 Low Voltage Audio Power Amplifier
- Crystal Oscillator 16MHz
- Secure Digital or (SD) card
- SD to Multi Media Card (MMC) adapter
- Liquid crystal display (LCD)
- Rail Connector

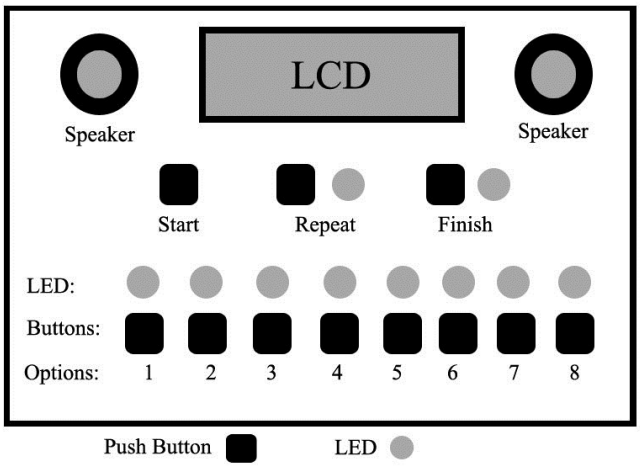


Figure 1: Block diagram

- Light Emitting Diodes (LEDs)
- Resistor (2.2 k ohm, 3.3 k ohm)
- Capacitor (0.1uF,10uF,10uF,100uF)
- Speaker
- Push Button

Developing in such way makes our device easily programmable and cheap to implement. It costs $2.5 per device on an average which is noticeably low. And, in case of SD card which can be increased according to the authority requirement, costs $1.5 for 1GB. The device contains this size of memory and system is designed to store the audio with the AAC file format, because it is more efficient than MP3. AAC makes a real difference at low bit rates, especially below 128kbps, but it don't make any practical difference at 256kbps and higher bit rates. This device is specially aimed for SEAR countries where authority needs more responsive, interactive and cheap way to make participate as many people as possible. Organizations which use this device can be able to decide the result on any query more accurately and precisely.

A. *User Interaction with the System*

User friendliness and system responsiveness are one of the key features of our system. A user who is intending to participate on our system can easily join. Question is always shown on LCD and push buttons of corresponding options blinked so that user can make sure for participating at first. When s/he becomes confirm to participate just need to press *START* button to play the audio from the beginning. The audio file is saved on SD card of our device. There can be multiple questions running by storing the corresponding audio files sequentially on SD storage. And every question directory is organized by four subdirectories and they are 'Question', 'Options', 'Instruction', 'Statistics' which store the pre-recorded audios of question, available options and corresponding LEDs with push buttons, necessary instruction and the updated statistics respectively.

While running the audio, user can play it again from the beginning if s/he thinks needed by pressing *REPEAT* button, otherwise the device waits until a threshold limit for getting the response from user. The threshold limit is set by default as 300 seconds which can be modified. At that time user decides the

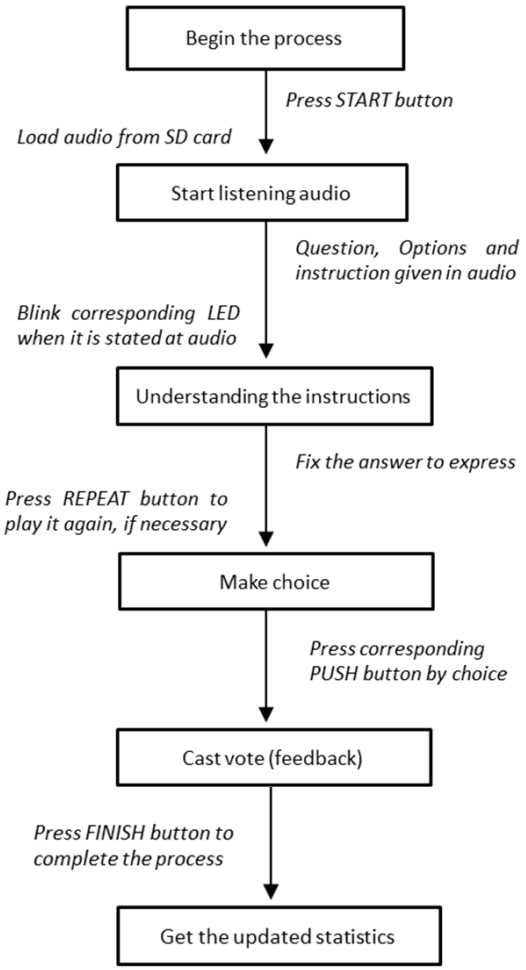


Figure 2: Device workflow

choice to enter by observing closely the blinking LEDs of options one after another. The first push button pressed by user, is considered as his/her opinion which gets stored on SD card and updates the statistics accordingly on system. After this, if user waits for the statistical audio update then also there is a threshold limit of 30 seconds to be passed. Otherwise if user is on hurry for the updated statistics then there is a *FINISH* which needs to be pressed for quick delivery from system. The whole process of device workflow is illustrated on figure 2.

B. Rapid Response on Disaster

A "disaster" generally includes any local hazard or act of terrorism which create a sudden requirement. In order to take any spontaneous action on any disastrous or hazardous situation, people need to be accumulated at first. And to do that understanding real picture in a big container is a must through reaching on majority based decision by learning mass opinion. Here our system plays a vital role by providing the mass participation platform and real-time statistical update. The system is capable of setting the questions in a priority based order which can be effective on any disastrous happening. For example, in flood season everybody is concerned about it, so flood related questions get highest priority to be prompted more than other questions. And getting the real-time statistics update one may find the most chosen options which may direct him/her self towards a safe direction what majority follows. Thus MassFeed's rapid response can be helpful for any misguided character and makes the system more convenient and effective.

C. Privacy Awareness in Information Processing

In a democratic region everybody have their own right to express his/her opinion. But, it is not as easy as it sounds. Terrorism, misuse of political privileges are the barriers to express the original voice of people instead of producing opinions under being influenced or threatened by someone. People get confused to deliver own opinion lest s/he face any further damage. So, anonymous participation can be a solution for expressing one's voice. In our system user doesn't need to provide his/her personal information to be a solid part of feedback taking process. To make the system widespread and increase the simplicity, we designed our system adaptive for any location and any type of question.

V. TEST RESULTS AND FUTURE WORKS

Currently we are evaluating our system in a few specific locations. We are incorporated with some volunteers from two universities of Bangladesh for testing our systems on every possible aspect. The responses have come till now are very positive and optimistic. People are enthusiastic about its features and implications than the web or mobile based methods. They accomplish the system with a complete score as it opens the door of being heard. Besides, the user-friendliness, real-time statistics update, user's anonymity are the main appreciated issues. In this phase of our system, the responses we got from mixed masses as it comes from both educated and illiterate people. We also hope to release MassFeed finally throughout several networks after our primary evaluation process.

Apart from the above mentioned features, MassFeed contains the capability to be incorporated by some data mining methodologies which is a primary plan of our future work. Through this approach a large amount of dataset will be analyzed to predict further or build up a meaningful decision. By learning public opinion, we also get information about the behavior and nature, identifying the trends of answers of a certain type of people of a location against different type of questions which also can be a research challenge to categorize the answers with some specific patterns. On top of that, covering the mass of a specific network significantly improve the social value and consciousness.

VI. CONCLUSION

In this paper, we propose a novel approach of capturing masses' opinions to make the decision impact more meaningful in our life. To our best knowledge, it is the first attempt on developing countries like Bangladesh to leverage the opportunity for the crowd to participate on any decision making phenomenon. Our system MassFeed facilitates its users

by providing a platform consisting user friendliness, highly accessibility and self-adaptation. As the system is audio based interactive approach it covers almost all kinds of people and scenarios to ensure the maximum participation. On top of that, we are working on extensive testing of our system in real world situation in conjunction with relevant mass audience and locations in order to increase its adaptability for make it more scalable and acceptable. Also, for those who are used to with mobile application usage, we are thinking of providing such application for sharing their opinion via voice mail or USSD (Unstructured Supplementary Service Data) communication with our device through GPRS. We expect that our system can bring a significant change in decision making situation of developing and under developed regions.

REFERENCES

[1] en.wikipedia.org/wiki/Atmel_AVR

[2] en.wikipedia.org/wiki/GSM

[3] http://en.wikipedia.org/wiki/List_of_countries_by_literacy_rate

[4] http://unpan1.un.org/intradoc/groups/public/documents/un/unpan048065.pdf

[5] http://help.instructure.com/entries/20913562-Polling-Audience-Response-Clickers

[6] www.scribd.com/doc/38170170/Cell-Phone-Based-Voting-System

[7] http://pewinternet.org/Reports/2012/Social-Vote-2012.aspx

[8] http://atmega32-avr.com/sms-based-voting-system-avr-gsm-project-using-atmega32-microcontroller/

[9] http://blog.nus.edu.sg/citations/2010/05/26/question-sms-%E2%80%93-sms-based-student-feedback-system/

[10] A. Gupta, W. Thies, E. Cutrell, and R. Balakrishnan, “mClerk: enabling mobile crowdsourcing in developing regions,” In CHI '12 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 1843–1852, 2012

[11] M. R. Dave, J. K. Singh, M. Tiwari and A. Khare, “Implementation of Intelligent Polling System Using GSM Mobile,” International Journal of Computer Technology and Electronics Engineering (IJCTEE) vol 1, issue 2, pp. 1843–1852, 2012

Parent-child Facebook Friendship: an Actuator for Generation Gap

Samiul Monir,*, Sabirat Rubya†,Hasan Shahid Ferdous‡
Department of Computer Science and Engineering,
Bangladesh University of Engineering and Technology, Dhaka-1000, Bangladesh
*samiulmonir@csebuet.org
†sabirat@csebuet.org
‡hsferdous@cse.buet.ac.bd

***Abstract*—Facebook is primarily designed to facilitate communication with different groups of friends, colleagues, and classmates. But recently with the increased craving for Facebook of parents/elder generations, we become interested to examine the influence of Facebook friendship among parents and their children on the rising intergenerational gap. A survey was conducted among Facebook users of different age range, genders, and customs. Findings reveal the type of information that kids cautiously hide from their parents and their impact on generation gap. The increased information gap is due to lack of mediation, transparency, mutual trust, intimacy, and equality. The results demonstrate the two-way effect, i.e., generation gap is particularly contributing to information hiding from parents and this opportunity of keeping secrets is again leading to a difference in values among younger and older generation.**

keywords - Child-parent behavior, Information, Social networking, Generation gap.

I. Introduction

The term 'generation gap', popularized in scholarly use during the late 1960s and early 1970s, refers to differences in values of older and younger generations [1]. It can also be described as certain psychological and emotional gap between parents or elder people and the younger ones. Swelling generation gap creates misunderstanding and lack of attachment between the parents and the children. Adolescents often feel individuating themselves successfully and creating separate identities. Therefore, they try to express opinions and internalize value systems that were distinct from their parents'. But parents usually do not perceive these salient dissimilarities between their own values and those of their offspring, in part because of their need to feel connected to their children. The fast paced development of technology has been considered as a key reason of the widening understanding gaps between the older and younger generations [2]. Children increasingly rely on personal technological devices and social networking media to define themselves and create social circles apart from their families, changing the way they communicate with their parents. In this paper, we are particularly interested in parent-children friendship in the context of the social networking site Facebook and its impact on generation gap.

Facebook, often called as the social media giant, was launched in February 2004 for Harvard students and in July 2008, more than half of Facebook users were outside universities [3]. According to the company's updated statistics, as of May 13, 2013, the site had a total of 1.1 billion active monthly users. Over 75% of parents who use Facebook are online 'friends' with their kids [4]. Facebook 'friends' are not necessarily the same as real life friends, but, as in real life, they are of different types. A key feature of the structure of Facebook is that, in general, information is available to all Facebook friends, including those who are family members. This raises issues of whether parents are welcome as Facebook friends by their children and what type of data they may hide after adding their parents as Facebook friends.

Many adolescents consider Facebook as their private lives or separate worlds, which their parents should not know about. To find out the effect of these limited Facebook friendships between a parent and a child, we have conducted an online survey that contained some general questions about the information hidden by the children from their parents. Statistics obtained from this questionnaire has revealed a two way relationship: generation gap is one of the reasons for information hiding and again, concept of the separate identities in these social networking sites like Facebook may exacerbate the generation gap among parents and children.

Digital media as a reason behind creating generation gap has been criticized by Lynn [5] as he argues that economically disadvantaged families experience the digital generation gap with particular intensity. He also points to the 'knowledge gap' that exists between those with digital expertise and those who lack it mentioning this fact as a source of shame. Further, an interview was carried out in [6] to examine the influence of Facebook on the intimacy level of parent-child relationship. Their study found out that the widespread usage of internet among adolescents have negatively affected the quality relationships between parents and children but Computer-Mediated Communication (CMC) can contribute to develop relationship by omitting the physical and social cues that are found in face-to-face interaction. Thus online communication like Facebook friendship can increase parent-child intimacy.

On the other hand, parents are rarely reported to be Facebook friends in [7] where sample of student users of Facebook have been interviewed. Older adults and parents are not usually welcomed even if they would be added. The majority of the reasons are related to embarrassment, social norms, and fear of punishment. In this paper, we aim to investigate these reasons

more practically and find out the attitudes of children towards the Facebook friendship with their parents and how these attitudes are leading to a generation gap.

In the following sections, we describe the methods we have adopted. The findings are discussed in the next section. We argue that in the context of friendships and potential friendships with older adults and particularly parents, a variety of concerns related to notions of privacy are raised, which militated against disclosing all information to them in Facebook. We also explain how different factors affected by this type of friendship can result in a rise in generation gap.

II. Methodology

To analyze how children remain cautious and react on their parents' activities, we prepared an online survey in July, 2012. We broadcasted our survey to different groups of students and professionals in Facebook. We also requested them to send us messages if they face any problem to complete the survey. We involved few of our foreign friends to participate and circulate this survey. Our prime target was to involve different ages of people from different customs. After removing a number of responses (incomplete submission, over age participant etc.) we received responses from 493 people.

A. Participants

Participants tended to be mostly in the age between 20 and 25 (63%) and 43% of them are male. But among teenager responders, approximately 56% are male participants. The average participant was 21 years old (SD = 3.4 years) and using facebook for at least 1.5 years (on average). 91% of the participants have their parents as Facebook friends whether parents are active or inactive. In maximum cases (98%), parents added participants after joining immediately and in 2% participants have added by themselves. 32% parents are well educated and 23% of them use their facebook account by themselves.

III. Findings

To better understand the influence of parent-child friendship in SNSs on generation gap, we divide the questions on the survey and show how the virtual friendship is leading to lack of some aspects needed to bridge the gap.

A. Mediation

One of the significant problems associated with having parents as Facebook friends relates to mediation [8]. Traditionally, an individual chooses what to disclose about his or her personal life, and how much information about his or her family should be known to friends. In real life, when these two parents and friends meet, it can be conjectured that this has been under the mediating influence and most likely watch of the individual in question. Parents being Facebook friends break this distinction. The individual's friends are able to see aspects of the family life and the parent is free to have contact with the individual's friends without the mediation of the individual. The user can thus lose control over the separation of these aspects of his or her life, facing an ongoing negotiation between sharing information with his or her friends while keeping information from his or her parents. To illustrate this fact, our survey enclosed three following questions that might reveal children's attitude towards the possible absence of mediation:

M1: Did you keep your parents in a different list?

-Yes (275, 55.8%)
-No (164, 33.3%)
-I did not accept them as Facebook friends (54, 10.9%)

M2: Do you think you have a different 'social life' in Facebook and your parents should not know about some of the events from that life?

-Yes (280, 56.8%)
-No (213, 43.2%)

M3: Do you think your parents may post some general stuff on your wall by which you feel embarrassed/ be teased among friends?

-Always (25, 5.1%)
-Sometimes (368, 74.64%)
-Never (100, 20.3%)

The development of the sense of hiding information about personal or social life starts when parents want their children to accept them as Facebook friends. Though majority of the respondents accept their parents' friend requests yet they keep them in different lists to limit their rambling across his profile and access all information. Furthermore, many of the responders believe that the separation among personal and social lives may be violated through Facebook friendship. About 57% of them do not want their parents to know about whatever aspects of their personal selves have been presented on Facebook whereas 62% of them are worried about the embarrassment from the general posts or comments (e.g., Did you have your breakfast, Bubla or How are you, love?) by their parents. Thus lack of the kids' presence as mediators leads to misunderstanding and cause more information gap. While a general tip for bridging the generation gap is not to do something that makes children ashamed of their parents. Such activities by parents raise the chance of their kids being teased.

B. Intimacy

In a previously discussed study, DeGroot, *et. al.* [9] wanted to know if relationship intimacy between parent and children could be predicted by online self-disclosure. The studies here appear to indicate that teens most often use social networking sites to connect with friends and build communities, something they are also doing offline. Nevertheless, risks lie in communication with dangerous strangers, lack of face-to-face

interaction, and the weakening of family ties. A large number of parents feel the amount of time their child spends on social networking sites interferes with family functioning and bonding.

As this generation uses an encrypted language understood by their peers exclusively, parents are less in tune with their children's communication; in essence, they are cut out. On the other hand, for some families, the availability of fast, non-verbal contact aids in keeping tabs on the kids. These issues are considered in our survey via the following two questions:

I1: How often do you communicate with your parents through Facebook?

-everyday (0, 0.0%)
-once in a week (25, 5.1%)
-once in a month (148, 30.0%)
-rarely (320, 64.9%)

I2: Do you use short forms of words like LOL, IDK etc. while communicating with your parents through Facebook?

-always (0, 0.0%)
-sometimes (240, 48.7%)
-never (253, 51.3%)

The answers reveal that, although adolescents spend a lot of time in Facebook [10] that was previously spent with family or real friends rather than with online friends, they rarely communicate with their parents. Moreover, teen children often use shortened form of words in their posts that seem to be 'jumble of characters' or 'jargons' to their elders [11]. The real feeling of communication is also greatly missed when a human voice is replaced by a buzz on their walls.

C. Power Struggle

As a child grows older, there will be an increasing desire by him to gain more power in his pursuit for greater independence and individuality [12]. When the child begins to form and strengthen his identity, a parent-child relationship is believed to grow closer if both parent and child are able to adjust their roles and correspondingly treat each other as equal individuals. But Facebook friendship between a parent and a child may hinder to achieve this independence. Parents get to know about many aspects of their children's social lives and they consider their kids to be liable to explain any unexpected events when asked. If the kids' explanations are not satisfactory then in some cases, they ban their kids from using Facebook as punishment. Besides, many parents allow their children to use Facebook only because they know their passwords and can keep an eye on all their activities. Children, on the other side, take it as interference against their individuality [13]. Our findings from the following questions give weight to these issues.

P1: Have you ever been asked concerning a matter that your parents knew about you from Facebook?

-Yes (114, 23.1%)
-No (379, 76.9%)

P2: Do your parents know your Facebook password?

-Yes (86, 17.4%)
-No (407, 82.6%)

Though a minor percentage of the respondents' answer was affirmative to the above two questions, the issue of the power struggle affecting generation gap is particularly important.

D. Transparency

Keeping certain information secret from the parents is another issue discussed as a reason behind generation gap. As many of the children keep their parents in different list after accepting them as Facebook friends, it is easily understood that they wish to hide some stuffs from them. The adolescents are worried of the fact that their parents may know about their private lives or any bad habits and thus the parent-child transparency is affected.

T1: Do you have a fear that your parents might know about any of your bad habits (e.g., smoking, drinking, etc.) by seeing your photos or through any comments/posts of your friends?

-Yes (178, 36.1%)
-No (186, 37.7%)
-I do not have such bad habits

T2: Do you have Facebook friends about whom you do not want your parents to know?

-Yes (299, 60.6%)
-No (194, 39.4%)

T3: What things do you hide from your parents?
The most selected option was 'Relationship status' and 'Photos and videos tagged by my friends'. It is evident that individuals remain cautious when they upload photos or videos as they know their parents are in Facebook. But they hide their photos in which their friends tag them because they don't want their parents to abruptly see something that they are not expected to see. This is also an effect of lack of mediation. Some reported that above 50% of information is hidden from their parents due to different reasons mentioned later.

The survey also included some general questions such as:
G1: Do you believe that the relationship with your parents would be more open/you would hide less info from your parents if Facebook did not exist?
G2: The main reason why you hide anything from your parents?

29% of the participants think that they would have to hide less information if Facebook did not exist whereas a high percentage (62%) of them mark 'Generation gap' as the reason behind information hiding. 'Fear of punishment' and 'Embarrassment' were chosen by 37% and 31%, respectively. Website like [14] is an alert proving the bulging generation gap resulting from parent-child friendship in social networking sites. Its slogan is "CONGRATULATIONS! YOUR PARENTS JUST JOINED FACEBOOK. YOUR LIFE IS OFFICIALLY OVER", where teens can share evidence of their parents' cluelessness. The site receives at least 20 submissions a day.

IV. Conclusion

We investigated how likely children response when their parents arrive on Facebook and what information they hide. Most of the participants of the conducted survey believe, it is the generation gap that keeps them away from their parents and make them keeping some secrets about their social lives. Whether they accept their parents as their Facebook friends or not, they remain cautious about separating their social lives from their personal lives. The lack of intimacy, transparency, and a possibility of power struggle arise when parents and children become Facebook friends, which ultimately direct to an increase in information gap between generations. One particular dimension to address in future would be the reasons why older adults, particularly mothers and fathers, 'friend' or consider 'friending' their children on and how to bridge the generation gap created from these virtual friendships.

References

[1] V. L. Bengston, "The Generation Gap.," *Youth and Society*, vol. 2, pp. 7–32, 1970.

[2] "Today's Generations Face New Communication Gaps." http://www.usatoday.com/money/jobcenter/workplace/communication/2002-11-15-communication-gap_x.htm.

[3] "Press Room, statistics [online] [Accessed 1 September, 2009]." http://www.facebook.com/press/info.php?statistics.

[4] "Given the Option, 30% of Teens Would Unfriend Parents on Facebook." http://tiny.cc/culwkw.

[5] L. S. Clark, "DIGITAL MEDIA AND THE GENERATION GAP," *Information, Communication and Society*, vol. 12, no. 3, pp. 388–407, 2011.

[6] V. H. hua Chen, T. Goh, and W. Li, "Welcome to Facebook: How Facebook Influences Parent-Child Relationship,"

[7] A. West, J. Lewis, and P. Currie, "Students Facebook friends: public and private spheres," *Journal of Youth Studies*, vol. 12, no. 6, pp. 615–627, 2009.

[8] S. Williams and L. Williams, "Space invaders: the negotiation of teenage boundaries through the mobile phone.," *Sociological review*, vol. 53, no. 2, pp. 314–331, 2005.

[9] J. DeGroot, A. Ledbetter, Y. Mao, J. Mazer, K. Meyer, and B. Swafford, "Attitudes Toward Online Social Connection and Self?Disclosure as Predictors of Facebook Communication and Relational Closeness.," *Communication Research*, vol. 38, no. 1, pp. 27–53, 2011.

[10] "Facebook Demographics Revisited 2011 Statistics." http://www.kenburbary.com/2011/03/facebook-demographics-revisited-2011-statistics-2/.

[11] "Has Texting Created a New Generation Gap Between Parents, Kids?." http://tiny.cc/1wlwkw.

[12] T. D. Golish, "Changes in closeness between adult children and their parents: A turning point analysis.," *Communication Reports*, vol. 13, no. 2, pp. 79–97, 2000.

[13] "48% of Parents Friend Their Kids on Facebook [STATS]." http://mashable.com/2010/05/03/facebook-parents/.

[14] "Oh Crap. My Parents Joined Facebook.." http://myparentsjoinedfacebook.com.

Incorporating Collaboration into the Design of Mobile Phones for the People with Low Literacy

Syed Ishtiaque Ahmed
Information Science, Cornell University
Ithaca, NY, USA 14850
sa738@cornell.edu

Maruf Zaber, Mehrab Bin Morshed,
Md. Habibullah Bin Ismail, Sharmin Afrose
Computer Science & Engineering, BUET
Dhaka 1000, Bangladesh
{maruf.zaber.09, mehrab.morshed, bahar61119,
sharmin.afrose}@gmail.com

***Abstract*—** The ubiquitous nature of the services of mobile phones has made it popular in most places in the world. Many operations of mobile phones often require the users to read and understand from the displays and keypads. This creates an obstacle for the people with low literacy (illiterate or semi-literate) toward getting the full services from their mobile phones. A number of technical solutions have been proposed in the literature, which involve audio or graphic assistances from the device to help the user. We extend this research but in a novel direction where we bring collaborative support from the social peers within the mobile phone interface to assist the users. Based on our six months of ethnographic fieldwork on a rickshawpuller community in Dhaka, we discovered the functions that they mostly use in their mobile phones and the way they do so. Instead of considering the use of mobile phone as an individual consumption of technology, we realized it as a social process of getting services where people collaborate with each other. Based on our observation, we propose a set of design and policy implications that would help them get better services from their mobile phones.

Keywords— Mobile Phones, Low Literacy, Collaboration.

I. INTRODUCTION

A large number of people around the world are illiterate. Moreover, there are many people who do not have the adequate literacy to operate electronic devices, which guide people through written commands. We put these cases apart from digital literacy, which is more related to people's inability to operate digital devices although they are capable of reading (and/or writing). In the developing world, many people are illiterate (unable to read, write, and understand short simple messages) and semi-literate (struggle to read, write, and understand short simple messages). We use the term "low-literate" to refer these two classes of people. With the burgeoning growth in the adaptation of mobile phones in those parts of the world, these low-literate people are now having mobile phones. Hence, many of the ICTD projects are being proposed based on the services through mobile phones. However, with a limited capacity to understand the written contents on the display, the low-literate users often struggle to get the benefit of these services.

A good number of research initiatives have been taken in recent few years to make the services of the mobile phone accessible to the low-literate people. If we consider the big picture, these research initiatives can be broadly classified into two big classes. One stem of research focuses on using effective graphical objects to help the users. The second big stem of research is focused on using speech or audio to help them. In section 2, we have discussed the previous research efforts in details. However, all those initiatives are more or less based on the device itself, and they considered the use of mobile phone mostly as an individual action. In our research, we have changed this perspectives based on our long field study over a rickshawpuller community at Dhaka. Instead of considering the use of mobile phone as an individual consumption of technology, we could realize this as a social process of obtaining services. We could understand how these people learn operations, share knowledge, collaborate each other, and thus help this technology sustain in the society. Based on that observation, we have proposed here a novel design paradigm where the interface of the mobile phones can allow the users to get connected to her social peers and get help from them.

After a thorough revision of the related previous works in section 2, we introduce our participants in section 3. This includes the introduction of the vehicle rickshaw, which is one of the prime means of short-range communication in Bangladesh. We also discuss the life of the people who pull rickshaws. In section 3, we discuss our research methods. We consider this crucial because getting into a community like this and collecting data was not very straightforward and we had to take a lot of decisions while being in the field. In the next section, we describe our findings on how the rickshawpullers takes the services from their mobile phones. Many of those findings matched with the similar sort of investigations by previous researchers. In the next section, we discuss our findings based on a focused group discussion with five of the rickshawpullers who were in our study. Based on their information, we could find out the people in their social networks, who provide them with support to consume various services of mobile phones. Later we present short description of those people to understand the nature of the help they provide with. Afterwards, we move to the discussion and design section, where we present our design implications based on our findings in the field. Finally, we conclude the paper with some future direction of research.

II. BACKGROUND

Chipchase et al [13] discovered from their study that turning on the mobile phones and accepting an incoming call are two most successfully executed operations on mobile phones by the illiterate people. However, dialing a number from the saved contact list was found to be challenging for them. Sending and reading text messages were not among their practices with their mobile phones. Later we have observed a number of initiatives by different commercial mobile phone companies where they came up with cheap mobile sets and dividing the contact list in a way that the illiterate users can easily use those. However, not much success on that initiative has been reported in the literature. It should be mentioned here that in most of the studies the users were more or less numerate, although they were wither illiterate or semi-literate.

A big stem of research in this area has been motivated by the idea that the people with low literacy would be more beneficial by the graphic contents than by the texts. So, lessening texts in the UI and incorporating graphical objects more effective has been studied in last few years. Medhi *et al* [4], [5], [6], [7] studied the effective use of graphical objects in the mobile phone interfaces for illiterate people and found out some interesting phenomena. One of the big challenges in using graphical objects is the absence of universally accepted picture for some particular functions. In many cases, low-literate people misinterpret the meanings of many pictures and that leads them to troubles. However, using hand-drawn icons instead of photos, and using moving icons instead of still pictures were found to be more effective in their studies. Number of initiatives later tried to improve the performance of the user interface either by adding graphical object to make better sense to the users, or by adding symbols to the keypad to help the users with a better management of contact list. However, the meaning of symbols and graphical objects are often culturally situated, and in many cases those vary from person to person even inside a same cultural territory. With the introduction of newer functions on mobile phones, newer symbols become necessary to convey the meaning to the low-literate users.

The second big stem of designing UI for illiterate people is motivated by the fact that audio commands require less literacy to understand [3], [6], [9], [14]. Audio-based UI hence has become an interesting area of research and a good number of studies have been conducted surrounding this idea. It has been found that the shorter audio commands are usually more effective than the longer ones. Also, the illiterate people often just perform according to the first command if there is more than one command to execute a particular function. Illiterate people were disturbed the illiterate-users although the semi-literate users got that helpful [13]. In a recent study, Medhi *et al* used visual demonstration of the healthcare issues to the illiterate users [4]. They found that the audio explanation helped the users to understand those clearly and without audio guidance they were more prone to commit mistakes.

In spite of a lot of promise, sound interface was found to be more error-prone in a longitudinal study in India, called “Avaaj Otalo” [14]. The users preferred typing keys instead of speaking in most of the cases. The speech recognizer was not robust enough to understand the users’ voice inputs in many cases. These limitations are imposing practical challenges in using sound/voice as the primary tool for the UI for the low-literate people. The variation of the colloquial often poses another problem while these interfaces are deployed in some low-literate communities.

Recently Knoche and Huang [1] studied the mobile using pattern of illiterate immigrants in Switzerland and found out how illiterates received and placed calls and manage contacts. They found that their interviewees considered text messages as an important asynchronous and cheap medium of communication. Later they developed an application [2] which allowed illiterates to listen to received SMS and compose text messages by augmenting words with touch-initiated text- to-speech support, icons for frequent phrases and by re-using words from previous messages. Lalji and Good [3] conducted a study on illiterate blue-collar workers from Mumbai. They adopted an incremental design strategy and founded that an interface that regulated the introduction of features may not be necessary. Medhi, Sagar and Toyama [4] designed UI for illiterates based on their ethnographic studies. They developed a job search application for domestic laborers. They also developed a generic map that could be used for navigating throughout a city. They [6] presented a quantitative evaluation of mobile interfaces that encompassed the full space of designs, spanning text, audio, and graphics. In another work [5] they illustrated user feedback using different representations with and without voice annotation. Medhi [7] also found that non-text designs are strongly preferred over text-based designs. Findlater, Balakrishnan and Toyama [8] explored how semiliterate users with very little education might benefit from a combination of text and audio as compared to illiterate and literate users. Sherewani, Palijo, Mirza, Ali and Rosenfeld [9] showed that a well-designed speech interface significantly outperforms a touch-tone equivalent for both low-literate and literate users. Joshi [10] designed an icon and color based visual phonebook for illiterates. Wiedenbeck [11] illustrated perceptions of usefulness of interfaces with icon-only and icon-label combination.

III. OUR WORK

A. Participants and Methods

For more than six months, we conducted ethnographic fieldwork with the rickshawpullers of a particular garage of Kamrangirchar. During this period of time, our study is divided into three phases. In the first phase we interviewed 10 rickshaw pullers in Kamrangirchar, a neighborhood south of capital city Dhaka. Ten Rickshaw pullers were picked randomly and were interviewed – all of them would work at that same garage. Half of our interviewees were single while rest of them had spouses and their family would reside with them in Kamrangirchar. Interviewees were picked carefully to cover all age range. Half of our interviewees were above 30. As we also interviewed the family members of the

rickshawpullers we got an insight of mobile phone using pattern of illiterate women and children too. The interviews were semi-structured and conducted in the native language of rickshawpullers for their ease. Our interview sessions were divided into five broad sections: a) introduction, b) biography of the participants, c) motivations behind using a mobile phone, d) usages of the memory of the mobile phones, and e) learning techniques.

In our second phase we invited five rickshawpullers from that garage to our university and had a focused group session in an interactive environment. One of the authors conducted the focused group discussion. This time we had a semi-structured discussion focused on the people they would take help from while using mobile phones. We figured out four basic nodes that help them to learn the operations - fellow rickshawpullers, the garage owner, nearby computer shop where they load media files to their cell phone, the repair shop and the nearby mobile phones' balance-recharging center.

In our third phase we revisited Kamrangirchar and talked to the people who would help the rickshawpullers learn and operate their mobile phones. We tried to understand the flow of information into the rickshawpullers of that garage elaborately.

IV. DESCRIPTION OF OUR STUDY

A. Rickshaw

Rickshaws are one of the popular means of transportation in Bangladesh. These are man-driven three-wheelers, having seats for two passengers behind the seat of the driver. However, in many cases more than two passengers are found to be on a single rickshaw. The drivers of the rickshaws are called 'Rickshawala' in Bangla. We would use the term, rickshawpullers here, which is an accepted English term for "rickshawala".

B. Kamrangirchar Rickshaw Garage

Kamrangirchar is a small area inside Dhaka district under the division of Dhaka. An estimated 400,000 people live in this 2.87 square kilometers area making it one the most densely populated areas in Dhaka Metropolitan. A 2006 survey [12] of slums in Bangladesh shows that the single largest concentration of slums in the Dhaka Metropolitan Area was in Kamrangirchar, and reported that of the approximately 300,000 people living there, 265,000 were slum dwellers. The area has 213 Rickshaw garages.

Figure 1. A rickshawpuller is pulling a rickshaw with passengers.

We picked one garage randomly. The garage had 73 Rickshaws and was managed by the owner of the garage and a manager. About 70 rickshawpullers worked in that garage. [Taka is the currency of Bangladesh. 80 Taka is approximately equivalent to 1 US Dollar at the time of the study]

V. USE OF MOBILE PHONES

From our individual interviews with the rickshawpullers, we came to know about the usage of their mobile phones. The rickshawpullers mentioned the following uses of their phones:

i) Talking: This is the most important and most frequent use of their mobile phones. They both place and receive calls to their friends, family members, and to other people. Sometimes it is important for their profession, too.

ii) Listening to Radio: Most of them love to listen to the programs broadcasted by the radio channels. Their mobile phones have that option of playing radio. They like the programs that have lots of songs from Bangla movies. Sometimes they keep on listening to the radios when they pull rickshaws.

iii) Watching Movies: One of the main attractions in their mobile phones is watching the movies. They usually watch Bangla movies, which they get installed in their mobile phones from the computer shops.

iv) Listening to Songs: rickshawpullers also install songs in their mobile phones at the computer shops. Then they listen to those songs whenever they want.

Figure 2. Rickshawpullers are taking rest at their garage.

v) Sending Money: Most of the rickshawpullers use their mobile phones to send money to their relatives living in the

rural parts of the country. Recently, a mobile phone based money-transferring service has been introduced in Bangladesh. Using that system anybody can send money to any other person having a person with a mobile network of a particular type.

vi) Taking Pictures: In occasional cases, they take pictures. Most of them do not know how to take pictures. The rest take pictures of their children or of themselves when they go to the wedding parties.

VI. INTERACTING WITH THE MEMORY

We were interested to understand how our participants stored, navigated, searched, and retrieved contents from their mobile phones. In the first round of the interview, we asked them how they would do that by themselves (without any help of other person). Many of these results match with the findings of Medhi et al. Hence our study reconfirms the behavior of low-literate people with the memory of mobile phones. However, in some cases we got some interesting new ways possibly invented by local people.

A. Contact Search

1) Memorizing Last 2/3 Digits:

Most of the participants were found to use this technique. They would memorize the image of the last 2/3 digits of the contact numbers and they were not aware of the fact that more than one contact could have the same last 2/3 digits.

2) Serial of the Contact:

Some of our participants would depend on others to get the contacts saved in their mobile phones. In this case, they would remember the position of the contacts in the contact list. They would do the same while searching a contact from their recent calling history.

3) Memorizing the Contacts' names as Images:

Some of the participants were found to use this strategy. They would memorize the contacts' names (saved by others) as images. They would try to remember how the letters look like when they were put together to produce the contacts' names.

4) Frequency of Letters

Some of the participants, instead of memorizing the whole name, would memorize the image of individual letters and would try to search contacts based on the frequency of a particular letter. For example: "ABBAS" has 2 "B"s in it.

B. Remembering Own Phone Number

1) Carrying a Paper With the Number:

Many of the participants were found to carry a paper with them, which had their phone numbers written on it.

2) Service from the Operators:

Some of the participants were found to use the service of the mobile operators for getting their own number. For example by pressing '*566#' they could see their own number on the screen.

3) Rhyming with 2 digits:

Some of them were found to memorize their own phone numbers just as a rhyme of pairs of numbers. "My number is, 'TwoTwo-OneOne-ThreeNine'", for example.

C. Searching Music and Videos

1) Thumbnails:

One of the easy ways for searching videos was to look at the thumbnails. Most of the participants whose mobile phones would support thumbnails, adopted this technique.

2) Serial of the Content

Once again the relative ordering of the media contents in one's mobile phone often helped one to recall any particular media content. Some of the participants would remember the serial of each of the media contents and thus access those.

3) Sequential Search:

Many of the participants were found to make exhaustive sequential search to find out the desired song or video, since they had no other way to do that. With the increase of the length of the list, this task becomes more and more difficult and time-consuming.

4) Memorizing the File Name:

In a few cases, they were found to memorize the file names (or last 2/3 digits) as an image to help search the media contents. This particular process was found to be similar to one of the techniques they would apply for searching a contact's number in the contact list.

VII. SOCIAL SUPPORT

Primary data didn't give us enough information to unfold how the community had saved numbers, or how they had installed the songs and movies in their mobile phones, etc. To get the answers of these questions, we invited five of them to our university to have a focused group discussion.

Most of them had some basic level of education and they were familiar with the English alphabet and numbers. They said they would type the numbers using the keypad, and then save it with the letter 'b'. The rationale was the name of our university starts with 'b'. This raised two questions in our mind: 1) How did they learn this mechanism

Figure 4. Five Rickshawpullers in a focused group discussion at the investigators' university.

of saving the number, and 2) What would they do if they had been told to save the phone numbers of more than one people of this same university.

In reply to our first question, they disclosed to us the fact that they learnt this from the owner of their garage. They garage owner saves the important numbers in their mobile phone whenever needed.

In reply to the second question, they told us that they would use a separate letter or symbol (such as, '*') to save the number of the second person of the same university though they told us that this technique was hard for them to remember if there were many contacts.

The rickshawpullers go to a local shop where CD/DVD, etc. are rented and sold to load songs or movies. They called it "computer shop". The person, who sits there in that shop, installs songs and movies in their mobile phones.

Our next question was regarding the money-transfer service of the network providers that the rickshawpullers use for sending money to their relatives in remote places followed by the set of questions on the repairing issues of their mobile phones. The two most common reasons for having problems in their mobile phones are: 1) the phones drop off their pockets when they bend for picking up something from the street; and 2) water enters into the phone when it rains.

We also asked them if they had any sort of collaboration among themselves in using mobile phones. Turned out, they had that a lot. The younger members of their community were more interested in checking different features of the mobile phones, and hence they knew better than the senior members.

To understand the collaboration in a better way of rickshawpuller community, we went to the people they mentioned to get help from.

A. The Garage Owner

Mr. G helps the rickshawpullers in his garage in using mobile phones. He mentioned four major reasons why rickshawpullers would need his help for their mobile phones: 1) when they need to save a new contact to their mobile phone, 2) when they need to play a song or a movie, 3) when they need suggestions for buying a new mobile phone, and 4) when their mobile phones do not function properly.

B. The Money-Transferring Shop

There were not this many services like transferring money through mobile phone at that time Mr. T started business which was 11 years back. Besides the monetary benefit, Mr. T feels moral obligation to help this people. He realizes how important it is for those people to send the money to their relatives. He always makes sure that their monies reach properly.

C. The Computer Shop

Mr. C is about 30 years old and he has been having this business for only 6 months. He sells CDs and DVDs of Bangla, Hindi, and English movies in his shop. At the same time, he loads these movies into the memory card of the mobile phones. Mr. C shows the rickshawpullers how to watch the movies or listen to the songs each time he loads those to their mobile phones. But he also receives complaints if they cannot play those later. In most of these cases, the customers make mistakes in operations.

D. The Repair Shop

The repair shop that the rickshawpullers under our study mostly use had one technician and two apprentices working under the owner, Mr. R. is also a technician himself. He has been living in Kamrangirchar since 1994. He took both formal trainings and apprenticed at different places. Besides repairing, Mr. R often gets requests from them to show how to use different features of the phone. He tries to teach them the operation of the functions in an easy way.

VIII. DUSCUSSION AND DESIGN IMPLICATIONS

From our field study, we understood that the interactions that the rickshawpullers make with their mobile phones should not be considered as their individual effort to break into the technology; rather it should be considered as a combined effort of their own understanding of letters and symbols, the help from their peers.

From our study and analysis, we make the following implications to the design of mobile phones' interfaces and interactions for the rickshawpuller community :

i) Bangla Fonts: In our study we found that the rickshawpullers are more familiar with Bangla letters and symbols, than the English ones.

ii) Graphical User Interface: We observed that the rickshawpullers are good at interpreting graphical objects. This particular observation supports the finding of Medhi et al. [5]. So, we propose to have a set of universal graphical code to guide them in the mobile interface.

iii) Screen Sharing: This is a novel concept in this context that we are going to propose here. Since we have found the rickshawpullers taking help from others for some

specific purposes, we propose a system that shares the mobile screen of the rickshawpuller to the person who is supposed to help them. The mobile networks, Wi-Fi, Internet, or Bluetooth could establish the connection between their phones.

Beside these design implications, we would bring forward some other relevant discussions that might introduce a bigger discussion around the technology and development issues in developing regions. First, from our field data, we could see that many of the rickshawpullers had their children going to primary schools though they don't get any training on using electrical devices. Second, we would expect governmental initiatives to help people in this regard. Experts might be sent to these communities once in a month to ensure their sound and healthy consumption of technologies. Third, we should think of mechanisms how to prevent over-charging for repairing phones, or loading digital contents into those. Establishing more shops and making more information available might help the users be aware of the actual prices for these services.

ACKNOWLEDGMENT

We express our gratitude to the owner of the garage where we conducted our study and Department of Computer Science and Engineering, Bangladesh University of Engineering.

CONCLUSION

This study reports the ways the rickshawpullers in one garage of Kamrangirchar use their mobile phones. These people could overcome the challenges of illiteracy by the power of their social network. This inspired us to propose incorporating collaboration in the interface of their mobile phones so that they can exploit the benefit of their social connections more effectively. A sound development of this proposed design and a robust user study would tell how effective the design could be had that been available. We put that part as our future work along with more detailed examination of the pattern of social relationships so that the whole idea of social user interface becomes effective to the low-literate communities.

REFERENCES

[1] Knoche, H. Huang, J., *Text is not the enemy - how illiterate people use their mobile phones.* In NUI workshop at CHI'12, March 2012.

[2] Friscira, E., Knoche, H., Huang, J. *Getting in touch with text: designing a mobile phone application for illiterate users to harness sms*. In Proceedings of the 2nd ACM Symposium on Computing for Development, ACM DEV '12, pages 5:1-5:10, New York, NY, USA, 2012. ACM.

[3] Lalji, Z., Good, J. *Designing new technologies for illiterate populations: A study in mobile phone interface design.* Interact. Comput., 20(6):574-586, Dec. 2008.

[4] Medhi, I., Sagar, A., Toyama, K. *Text-free user interfaces for illiterate and semiliterate users. Information* Technologies *for* International *Development*, 4(1):37-50, Oct. 2007.

[5] Medhi, I., Patnaik, S., Brunskill, E., Gautama, S. N., Thies, W., and Toyama. K., *Designing mobile interfaces for novice and low-literacy users* ACM Transactions on Computer- Human Interaction, *18(1):2:1-2:28, May 2011.*

[6] Medhi, I., Prasad, A., Toyama, K. *Optimal audio-visual representations for illiterate users of computers.* In Proceedings of the 16th international conference on World Wide Web, WWW '07, pages 873-882, New York, NY, USA, 2007. ACM.

[7] Medhi, I., Gautama, S. N., Toyama, K. *A comparison of mobile money-transfer uis for non-literate and semi-literate users.* In Proceedings of the 27th international conference on Human factors in computing systems, CHI '09, pages 1741-1750, New York, NY, USA, 2009. ACM.

[8] Findlater, L., Balakrishnan, R., Toyama, K. *Comparing semiliterate and illiterate users' ability to transition from audio+text to text-only interaction.* In Proceedings of the 27th international conference on Human factors in computing systems, CHI '09, pages 1751-1760, New York, NY, USA, 2009. ACM.

[9] Sherwani, J., Palijo, S., Mirza, S., Ahmed, T., Ali, N., Rosenfeld, R. *Speech vs. touch-tone: telephony interfaces for information access by low literate users*. In Proceedings of the 3rd international conference on Information and communication technologies and development, ICTD'09, pages 447-457, Piscataway, NJ, USA, 2009. IEEE Press.

[10] Joshi, A., Welankar, N., BL, N., Kanitkar, N., Sheikh, R. *Rangoli: a visual phonebook for low-literate users.* In Proceedings of the 10th international conference on Human computer interaction with mobile devices and services, MobileHCI '08, pages 217-223, New York, NY, USA, 2008. ACM.

[11] Wiedenbeck, S.. *The use of icons and labels in an end user application program: An empirical study of learning and retention.* Behaviour Information Technology, 18(2):68-82, 1999.

[12] Measure Evaluation / NIPORT (2006). *Slums of urban Bangladesh: mapping and census 2005.* Centre for Urban Studies / Measure Evaluation / National Institute of Population Research and Traning. Accessed 9 June 2007.

[13] Findlater, L., Balakrishnan, R., and Toyama, K. *Comparing semiliterate and illiterate users' ability to transition from audio+text to text-only interaction*. Proceedings of the 27th international conference on Human factors in computing systems, ACM (2009), 1751–1760.

[14] Patel, N., Chittamuru, D., Jain, A., Dave, P., and Parikh, T.S. *Avaaj Otalo—A Field Study of an Interactive Voice Forum for Small Farmers in Rural India*. Proceedings of the Proceedings of the 28th international conference on Human factors in computing systems (Atlanta, GA, USA, 2)

SNS Search in Under-Developed Countries - Survey and Statistical Analysis

Saif Ahmed[1], Md. Tanvir Alam Anik[1], Mashrura Tasnim[1], Hasan Shahid Ferdous[2]
[1]Department of Computer Science and Engineering, Bangladesh University of Engineering and Technology
[2]Department of Computing and Information Systems, Unversity of Melbourne, Australia
Email: saif.ahmed@csebuet.org, tanviranik@gmail.com, mashrura_cse@yahoo.com, hasan.shahid@csebuet.org

***Abstract*—The last decade has seen the emergence of the Social Networking Sites (SNS) and researchers are investigating the useful applications of this technology in various areas apart from its recreational value. Its ubiquitous presence has enabled people to obtain customized information seamlessly from their acquaintance. There have been many works that analyzed the types and topics of questions people ask in these networks and why. Topics like what motivate people to answer such queries, how to integrate the traditional search engines and SNS together are also well investigated. In this paper, we focus on the use of this technology in under-developed parts of the world and the new doors it has opened for its inhabitants. Based on a structured survey on 328 persons, we have tried to emphasize the differences between social search and traditional web-search. Our statistical analysis finds the correlations among different relevant parameters and provides insight that one might require to consider while developing any application for SNS based searching.**

Keywords - Social Networking Sites, Search, Survey, Information.

I. Introduction

Individuals encounter a wide range of information needs in their everyday life. These include questions or queries, recommendations about career development, factual knowledge regarding sophisticated technologies, rhetorical thoughts of life events, opinions about a major purchase, etc. Despite of the digital revolution over the past three decades that has enabled us to store and maintain large collection of data in a tiny amount of space and the evolution of search engines (SE) that has made it possible for us to look into tremendous amount of information within seconds, the recent developments in social networking sites have opened a new dimension in information searching. Recent trends have shown that many people tend to turn to their social networks for information rather than using the formal searching tools. Social network sites (SNS) are connecting individuals to one another with whom they have a previously established offline connections or different degrees of relational closeness in online or in real life. Thus it is likely that people turn to SNS as an efficient way to tap these connections for information-seeking purposes.

In this study, we will use Facebook as an example SNS without losing any generality and seek to understand the extent to which Facebook users are likely to use the site for their information-seeking as well as information-retrieval activities. With one billion monthly active users and more than half a billion daily active users [1], currently (as of 14th April, 2013) it is the number two site in the world considering Internet traffic, according to Alexa ranking [2]. On an average, the users spend 10.5 billion minutes per day on it, make 421 million status message posts, 3.2 billion likes and comments, and have 140 friends in their network [3]. The average age of SNS users has also increased in recent years: among American Internet users, 70% of 30–49 year-olds, 51% of 50–64 year-olds, and 33% of those 65 or older now have a profile on a SNS.

SNS provide users with source of information that is complementary to that provided by search engines. Search engines provide information that comes from ubiquitous source, i.e., web, in contrast to SNS that can provide objective data from a variety of sources on a variety of topics and is highly tailored to an individual. Information obtained from SNS is also found to be highly trusted, as we know the individual behind the information too. We can simply ask the members of our social network and get useful information that the researchers found quick, useful, and in many cases, more robust.

For better understanding, we will emphasize on individuals' question and answer exchange in their social networks. We will primarily focus on the developing regions of the world and see how SNS search has made significant changes in the way people access information here. We have conducted a survey of 328 people using Facebook. Our survey covered topics such as the types and topics of the questions asked, frequency of getting responses, reasonable response time of getting appropriate responses, frequency of questions asked, and motivations behind question asking on social networks.

We discuss the related works in Section II. Section III presents the changes that SNS search is bringing in the under developed parts of the world to eradicate the digital divide. We present our survey data and analysis with its implication in Section IV before concluding.

II. Related works

Lampe *et. al.* [4] analyzed how the use of Facebook has changed over time, using three consecutive years of survey data and thorough interviews with a few of the survey people. They reported that though the uses of the site remain relatively constant over time, but the perceived audience for user profiles and attitudes about the site showed differences over the study period. They find that patterns of use, perception, and attitude somewhat changed over the time. Their study, consistent with

others, found that the number of friends and time spent on Facebook increased at first and then leveled off.

One of the important studies in SNS based information search is done by Efron *et. al.* [5], who identified that micro-blogging services like www.twitter.com are gradually becoming a popular venue for informal information interaction. They showed that the act of asking questions in Twitter is not analogous to information seeking in more traditional information retrieval environments, which contextualize these articulations through analysis of a large body of tweets.

Teevan *et. al.* [6] discussed the types of information people used twitter to find, for example, breaking news, real-time content, popular trends, etc. This paper presented the systematic overview of search behavior on Twitter and differences with web search using questionnaire data along with analysis on query logs. They found that Twitter results included more social content and events, while web results contained more facts and navigation. Based on their study, they recommended that search engines can use trending Twitter queries to discover additional queries that have strong temporal components.

Lampe *et. al.* [7] investigated the Facebook user characteristics based on a survey of 614 people who used it to ask something. They identified the perception of the relationships within network members as significant predictors of information seeking approach. They did not show any comparison between SNS and SE regarding obtaining any particular type of information. This question is addressed by Morris *et. al.* [8]. They explored the pros and cons of using SNS as information source and compared user interaction when they search anything either on SNS or SE, involving 12 participants on their study. They find that 53% of the users received quick responses from SNS and 83% received responses eventually as well.

The type of questions and answers in SNS are investigated by Morris *et. al.* [9] using a study of 624 people about their Facebook usage experience. They also explored the relationships between answer speed and quality, properties of participants (age, gender, and social network usage habits), and their questions (type, topic, and phrasing). Their study complies with the findings of many other researchers that while traditional SE is good for objective queries, SNS shows better results and interactions for subjective queries. There are many motivations for asking questions in SNS - among them the most important reason was that people in our social network knows our *context* better, therefore may provide more relevant answers. Often people turn to SNS regarding objective questions if knowing the answer is not urgent, in the hope that some other friend in his network already knows the answer and will share his knowledge with him in due time.

Panovich *et. al.* [10] evaluated the role of *tie strength* in question-response behavior as an indication of how close the relationship is – close friends are strong ties, while acquaintances are weak ties. In their study, they asked 19 participants to ask some technological recommendation questions through status messages. After the participants rated the received answers' quality, they compared that with a tie

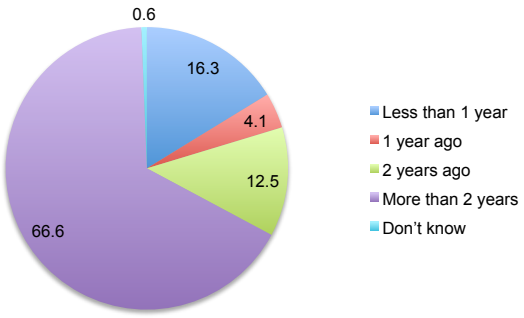


Fig. 1: Percentage of Facebook usage period.

strength metric, and found that stronger tie provides better answers than weaker ties, in general. Also, they found that friends who have expertise in the question topic provide more trustworthy answer irrespective of strong or weak ties.

Farnham *et. al.* [11] studied the suitability of *So.cl*: a web application that combines web browsing, search, and social networking, designed for the purposes of sharing and learning around topics of interest by taking feedback from 32 college students. Their findings present the importance of social media for inspiring learning around new topics through social connections. They found the easy, lightweight integration of sharing around search in *So.cl* effectively fostered serendipitous, informal learning online.

None of these researches investigate the difference in question-answer behavior in different parts of the world. Yang *et. al.* [12] addressed this issue and identified some key differences between SNS search in the Western and Eastern cultural hemisphere. Their survey included people from US and UK representing the Western culture and people from China and India representing the Eastern culture. They concluded that people in the Eastern culture are somewhat more likely to use SNS for getting objective information than their counterpart and use it more often for the purpose. They explained this phenomenon using existing and established knowledge from sociology study that Western cultures are associated with an analytic and low-context cognitive pattern, along with individualism, while Asian cultures are associated with a holistic, high-context cognitive pattern, along with interdependence and collectivist social orientation. Our initial findings match with them, except they did not include another possible explanation of this behavior - the existing web infrastructure deficit in the developing and undeveloped countries, commonly known as the *Digital Divide*. In our work, we will elaborate on this explanation.

III. Unite the Digital Division by Means of SNS

The term *Digital Division* indicates the difference in technological advancement between the developed and developing parts of the world. Availability of all sorts of computing devices in the developed region, especially the recent explosion in the smart-phone usage has made their web presence ubiquitous nowadays. Traditional search engines are very effective in capturing the required information regarding any

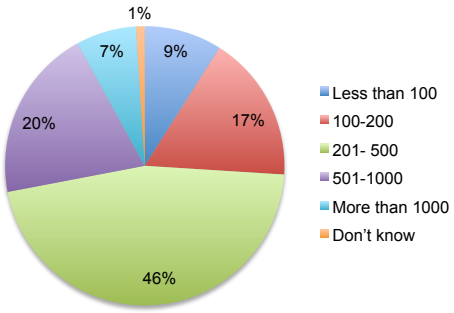


Fig. 2: Number of Facebook friends.

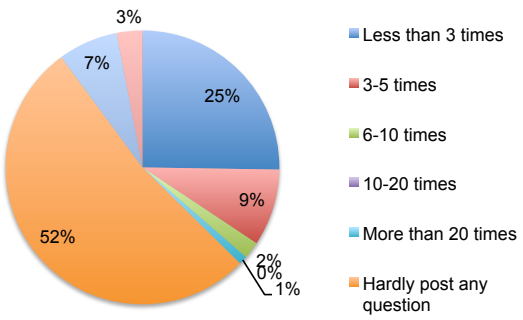


Fig. 4: Frequency of posting query in Facebook (per week).

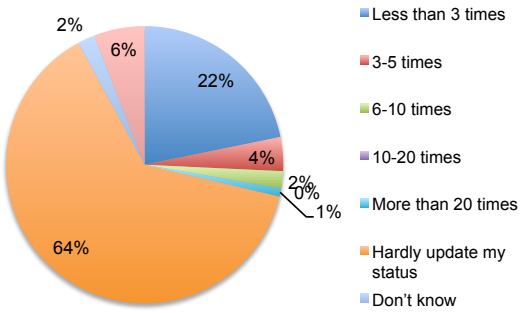


Fig. 3: Facebook status update statistics (per week).

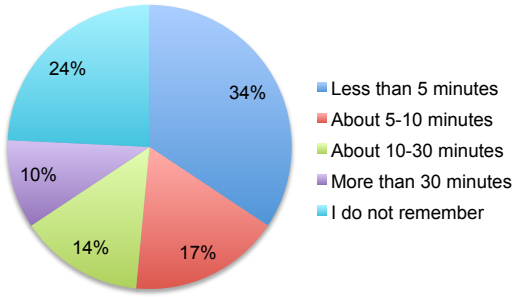


Fig. 5: First response time for the queries.

government or business services for them as it is already there in Internet.

The scenario is quite opposite in the other parts of the world where the web culture has not flourished yet. We may focus on the South-East Asia region as an example of the developing part of the world. Scarce accessibility to Internet and unavailability of updated information there create twofold problem in information retrieval in this region. First, search engines cannot provide satisfactory response of queries, as the information is not there in the web. Second, people are not used to search their required piece of information even if it is available in Internet. Though the Governments in these countries are trying to eradicate this digital division, it is proved as not easy. The world remains 'divided' and probably will remain so for a long time from now unless some drastic measure is not taken.

To show evidence on how SNS can pave a way to end this 'digital divide', we consider 'Bangladesh' as a representative country from the South-East Asia. Bangladesh is ranked 3rd among the 8 countries in this region in the e-Government ranking. Internet access is available to only 5 percent of her citizens and many of those who have access to Internet use it seldom. But if we consider the SNS presence of the people in Bangladesh, they are not far behind [13], [14]. Roughly 43% of the Internet users in Bangladesh use Facebook, proving that a significant part of our Internet users are SNS user too. Though SE cannot give them the data that is not there in the web, through SNS, their query can reach hundreds of the people of their acquaintance. Perhaps this is not the end of 'digital division' mentioned earlier, but we are getting a bit closer to unify the world in terms of information searching and retrieval capacity.

IV. Survey Results

We have conducted a structured survey on 328 people. Most of them are students from different universities and other reputed educational institutions in Bangladesh. Almost all of them have Internet access to various extents, and with negligible exceptions, all of them are SNS users. The results show interesting patterns in information retrieval experience through SNS. We have conducted both online and offline survey using the same questionnaire in English, which be found in [15].

A. Demographic Description of the Participants

We have collected 328 responses in total, all of which are undergraduate level students. 93 percent of our participants are from the age group of 18–24 years. The male-female ratio was not equal, 78 percent being male. This is the common and expected ratio of male and female students in the undergraduate level in Bangladesh. 98 percent of our survey participants were unmarried. We will see if these demographic properties have any impact on their Facebook usage pattern in later parts of this paper.

B. Question Asking Pattern through Facebook

Almost all of our participants (99.1%) have their own Facebook account and majority of them are using it for more than two years (Fig. 1). Majority of them have about 201–500 friends in Facebook (Fig. 2). When asked about how frequently

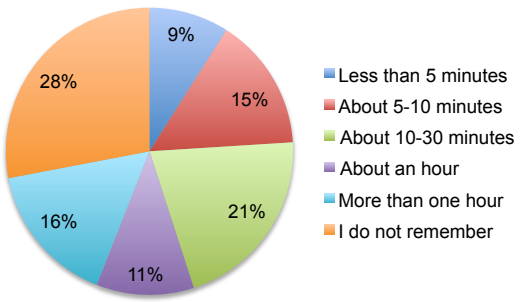


Fig. 6: Satisfactory response time for the queries.

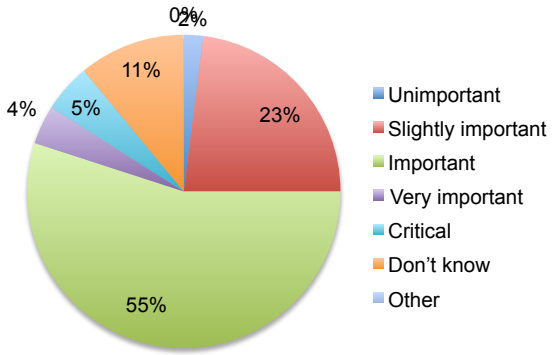


Fig. 7: Rating of the responses received.

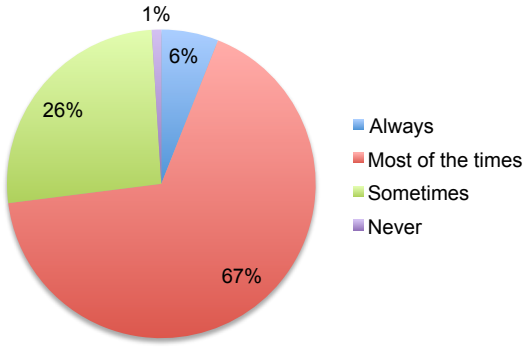


Fig. 8: Ratio of satisfactory response.

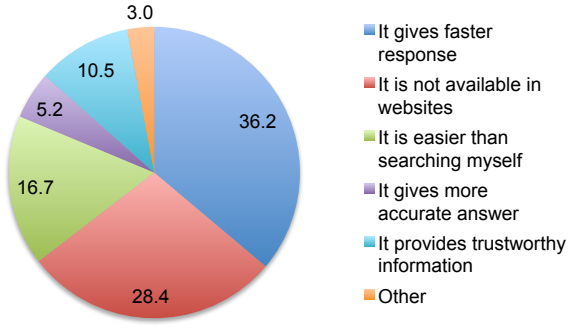


Fig. 9: Reason behind using Q/A in Facebook.

they update their Facebook status (Fig. 3), most of them said they hardly update their status (64%) or less than 3 times per week (22%). 73.8 percent of our participants have asked some questions or opinions through Facebook, 23.5 percent have never used Facebook for the purpose and 2.7 percent of the participants did not reply to that question.

Fig. 4 shows that 52% participants hardly post any question through Facebook and 25% of them posts less than three questions in a week. Though this scenario depicts that searching for information is not part of everyday use of SNS, the response time showed in Fig. 5 and Fig. 6 prove that SNS can be an influential source of information. More than 50% of the participants receive first response for their query and 45% of them receive satisfactory answer within 30 minutes. About 67% of the people has provided opinion that they get satisfactory response to their queries from friends most of the times, if not always and 26% people get it 'sometimes' (Fig. 8).

Technology, entertainment, and current events (Fig. 12) appear to be as most popular topic of asking questions. About 67% of the people has provided opinion that they get satisfactory response to their queries from friends most of the times, if not always and 26% people get it 'sometimes'. Again, majority of the participants (55%) rate the answer they get from their friends as important (Fig. 7). 52% of the participants said that they also have obtained valuable information many times from the questions other people posted, whereas 35% have said 'a few times' for the same. Getting faster response and unavailability of information on web were reported as two major reasons behind asking questions (Fig. 9).

47% of participants said that the results obtained from SNS and SE might differ to some extents and SNS can add some additional information or perspective to the problem at hand. Majority of the people (56%) will choose Google for information searching, but 40% said that they would decide either Facebook or Google depending on the query under consideration. In case they have limited time or bandwidth, 67% will use Google, 27% will decide depending on the query, and only 5% will use Facebook solely. If they do not get satisfactory information from Facebook, most of them (46%) will ask someone personally about it, almost a similar percentage of people will search using traditional search engines (42%) (Fig. 10).

C. *Answering Behavior*

Most of the people response to a query they see, at least 'sometimes' (47%) (Fig. 11). The main motivation being helpful to others (Fig. 13). To assist others in their friend list, 49% of the people often do a Google themselves to provide an answer to a question posted in Facebook, while 24% said that they ask another friend personally to know the answer and let the asker know about it.

D. *Inter-relation among the Parameters*

We used chi-square test of independence to find out interdependencies among different parameters based on demographics and SNS usage pattern (Table I). Most of the tests

TABLE I: Chi-Square test of independence on SNS question asking and answering behavior

Null Hypothesis	Independent of -	Chi-square value	degree of freedom	Dependent (Yes/No)	Comment	Z score
Type of Question	Time to get first response	46.7	36	No		-1.23
	Time to get satisfactory response	40.8	45	No		0.39
Gender	Time to get first response	1.59	4	No		0.88
	Time to get satisfactory response	6.6	5	No		-0.67
	Type of question asked	3.88	7	No		0.82
Length of Facebook use	Time to get first response	14.7	16	No		0.12
	Time to get satisfactory response	26.5	20	No		-1.04
	Importance of Facebook as information source	19.5	24	No		0.60
	Number of query update in search engine	46.5	16	Yes	People who uses Facebook longer are likely to modify their query and search again	-3.77
	Preferred source for information search	33.3	20	Yes	People who have been using Facebook for 'More than 2 years ago' shows slightly more inclination to use Facebook as an information source than others	-1.86
	Type of questions asked	22.4	28	No		0.71
Number of friends	Time to get first response	28.4	16	Yes	First response time decreases slightly with the increase in number of friends	-1.91
	Time to get satisfactory response	24.1	20	No		-0.71
	Importance of Facebook as information source	26.4	24	No		-0.43
	Number of query update in search	40.3	20	Yes	People who have more friends modify their query more than others	-2.60
	Preferred source for information search	129	25	Yes	Though Google is the first choice as an information source for all groups, people with moderate number friends (201–500) chooses it more than others	-6.0
Frequency of status update	Time to get first response	24.5	24	No		-0.17
	Time to get satisfactory response	28.3	30	No		0.14
	Importance of Facebook as information source	50.6	36	No		-1.61
Frequency of posting questions	Time to get first response	48.2	24	Yes	Frequent askers get response quickly	-2.82
	Time to get satisfactory response	26.6	18	No		-1.36
	Importance of Facebook as information source	22.7	24	No		0.09
	Type of questions asked	59.1	42	Yes	People who 'Hardly post any question' asks about 'current event' or 'entertainment' more than others	1.73
Response to others' questions	Importance of Facebook as information source	14.4	24	No		1.53
	Preferred source for information search	45.4	20	Yes	People who always reply to others choose Facebook, but who 'sometimes' response to others choose Google more	-3.1

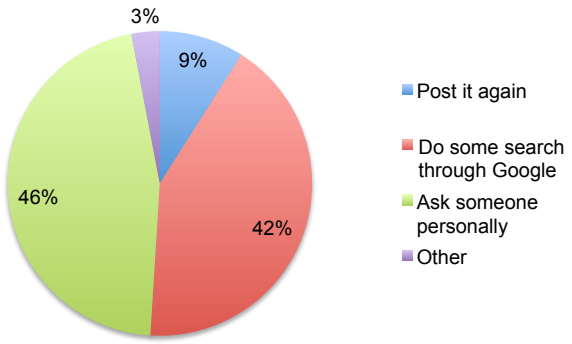


Fig. 10: Actions when no satisfactory answer is obtained from Facebook.

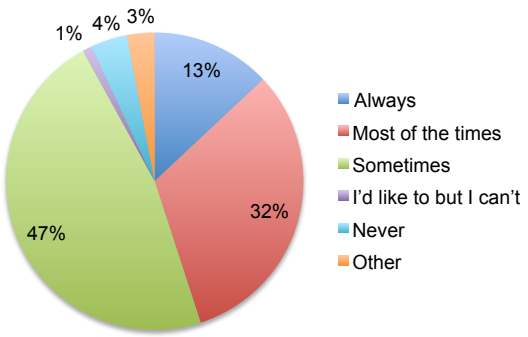


Fig. 11: Response to others' queries.

showed independent behavior, for $p < 0.05$. When the test results showed dependency, it means, at least one of the samples deviate significantly from the other samples. The test does not identify where the differences occur or how many differences actually occur, so we calculated the Z-scores of different samples for that query to see the dependencies or anomalies among them.

Our analysis shows that though the gender ratio among the participants is not equal (78/22), their Facebook usage patterns, specially the first/reasonable response time are independent of their gender. And though the lifetime of the users'

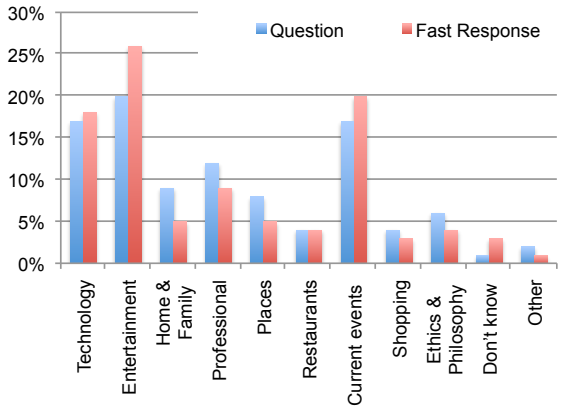


Fig. 12: Question topic and response preference.

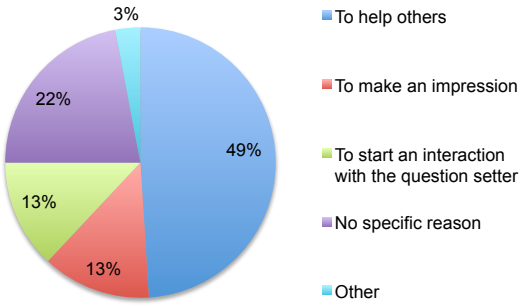


Fig. 13: Motivation behind response to others' queries.

Facebook account has no co-relation with response time and number of times they repost their queries, it is co-related with the information source they use for asking queries. People who have used Facebook for more than two years have considered using 'both Facebook and Google, depending on the query' more than any other sample populations.

The length of Facebook usage or the number of friends has no impact on the topics of their posted questions. However, the more friends they have, the quicker the first response to their queries comes. But interestingly, the reasonable response time is independent of the number of Facebook friends. Also, according to our analysis, those who have more friends, prefer to use 'both Facebook and Google' as an information source more than the other sample groups.

The analysis shows that there is no apparent relation between the frequency of normal status update with first/reasonable response time, preferred source of information search, or question topics. However, frequency of question posting has some dependency with the first response time and question topics - people who posts question frequently gets their first response quicker than others and people who seldom posts queries are more interested in 'current events' and 'entertainments'. We also found that the topic of question has no impact on first/reasonable response time or preferred source of information. Those who responses to others' queries 'most of the times' or 'sometimes' choose 'both Facebook and Google' as their preferred information source.

V. Conclusion

SNS search is gaining popularity on its own merit, and researchers around the world are considering it as a big change in the way we access or search information. Our research has emphasized its differences and importance with respect to the under-developed parts of the world, focusing on the digital-gap it has with the developed regions. With our background studies and structured survey, we have identified the types and topics that draws significant interest, first and reasonable response times, preferred information source, etc. Our statistical analysis could show inter-relation among these parameters and provides insight into the system. We are now working on developing a Facebook app using these information that will assist the users to obtain real life information from Facebook in a better optimized way than now.

References

[1] "Key Facts - Facebook Newsroom," http://newsroom.fb.com/Key-Facts.
[2] "Alexa Ranking of Facebook," http://www.alexa.com/siteinfo/facebook.com.
[3] "Facebook Usage Statistics," http://www.thesocialskinny.com/100-social-media-statistics-for-2012/.
[4] C. Lampe, N. B. Ellison, and C. Steinfield, "Changes in use and perception of facebook," in *Proceedings of the 2008 ACM conference on Computer supported cooperative work*, ser. CSCW '08. New York, NY, USA: ACM, 2008, pp. 721–730. [Online]. Available: http://doi.acm.org/10.1145/1460563.1460675
[5] M. Efron and M. Winget, "Questions are content: a taxonomy of questions in a microblogging environment," in *Proceedings of the 73rd ASIS&T Annual Meeting on Navigating Streams in an Information Ecosystem - Volume 47*, ser. ASIS&T '10. Silver Springs, MD, USA: American Society for Information Science, 2010, pp. 27:1–27:10. [Online]. Available: http://dl.acm.org/citation.cfm?id=1920331.1920371
[6] J. Teevan, D. Ramage, and M. R. Morris, "#twittersearch: a comparison of microblog search and web search," in *Proceedings of the fourth ACM international conference on Web search and data mining*, ser. WSDM '11. New York, NY, USA: ACM, 2011. [Online]. Available: http://doi.acm.org/10.1145/1935826.1935842
[7] C. Lampe, J. Vitak, R. Gray, and N. Ellison, "Perceptions of facebook's value as an information source," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '12. New York, NY, USA: ACM, 2012, pp. 3195–3204. [Online]. Available: http://doi.acm.org/10.1145/2207676.2208739
[8] M. R. Morris, J. Teevan, and K. Panovich, "A comparison of information seeking using search engines and social networks," in *ICWSM*, 2010.
[9] ——, "What do people ask their social networks, and why?: a survey study of status message q&a behavior," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '10. New York, NY, USA: ACM, 2010, pp. 1739–1748. [Online]. Available: http://doi.acm.org/10.1145/1753326.1753587
[10] K. Panovich, R. Miller, and D. Karger, "Tie strength in question & answer on social network sites," in *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, ser. CSCW '12. New York, NY, USA: ACM, 2012, pp. 1057–1066. [Online]. Available: http://doi.acm.org/10.1145/2145204.2145361
[11] S. Farnham, M. Lahav, D. Raskino, L. Cheng, and T. Laird-McConnell, "So.cl: An interest network for informal learning," in *International AAAI Conference on Weblogs and Social Media*, 2012.
[12] J. Yang, M. R. Morris, J. Teevan, L. A. Adamic, and M. S. Ackerman, "Culture matters: A survey study of social q&a behavior," in *ICWSM*, 2011.
[13] "Internet User Statistics," http://data.worldbank.org/indicator/IT.NET.USER/countries.
[14] "Facebook User Statistics," http://www.socialbakers.com/facebook-statistics/page-2/?interval=last-6-months.
[15] "Survey questionnaire on Q/A behavior in Facebook," https://docs.google.com/forms/d/1Px8hXj9NEz1SuO9skaBlmHKnUW9X13gkmHsrzXB4eQg/viewform?pli=1.