DEVELOPMENT OF SOFTWARE FOR THE BASIC LINE-OF-SIGHT PARAMETERS CALCULATION

***Abidur Rahaman*[1] , *Md. Shariful Alam*[2], *ZH Mozumder*[2]**

[1]Dept.of Information and Communication Engineerimg,Noakhali Science and Technology University,Noakhali,Bangladesh

[2]Dept.of Electrical and Electronic Engineering,University of Dhaka,Dhaka,Bangladesh

{*abidur@nstu.edu.bd, zmozumder@du.ac.bd*}

***Abstract**: In this paper we have developed a software by which the general parameter of a line-of-sight (LOS) microwave link can be calculated. We have also put here an overall concept about the link parameters. A Line-of-sight microwave link is designed. For a LOS link implementation there are two steps to follow: at first, we have to make a survey to collect a few data and then we have to calculate some data with the help of survey data. The link parameters which are crucial to the design have been calculated. All important parameters like Fresnel zone, fade margin, effective earth curvature, antenna tower height and the minimum transmitter power for a given BER have been calculated. In the link budget calculation all of the losses like fading loss, absorption loss, feeding loss and noise figure of the receiver are considered. Finally, a computer GUI program has been developed for the enhancement of a complete usable LOS software which original coding was done in C language to be used by any designer.*

***Key Words**— Microwave, Line of Sight (LOS), Link parameter, LOS software.*

I. INTRODUCTION

Microwaves are electromagnetic waves with wavelengths ranging from 1 mm to 1 m, or frequencies between 0.3 GHz and 300 GHz. The term microwave refers to electromagnetic energy having a frequency higher than 1 Gigahertz (billions of cycles per second), corresponding to wavelength shorter than 30 centimeters. The microwave range includes Ultra-High Frequency (UHF) (0.3–3 GHz), Super High Frequency (SHF) (3–30 GHz), and Extremely High Frequency (EHF) (30–300 GHz) signals.

A Line-of-Sight microwave link uses highly directional transmitting and receiving antennas to communicate via a narrowly focused radio beam. The transmission path of a Line-of-Sight microwave link can be established between two land-based antennas, between a land-based antenna and a satellite-based antenna, or between two satellite antennas.

A link budget is the accounting of all of the gains and losses from the transmitter, through the medium (free space, cable, waveguide, fiber, etc.) to the receiver in a telecommunication system. It accounts for the attenuation of the transmitted signal due to propagation, as well as the antenna gains, feed line and miscellaneous losses. A simple link budget equation looks like this:

Received Power (dBm) = Transmitted Power (dBm) + Gains (dB) − Losses (dB).

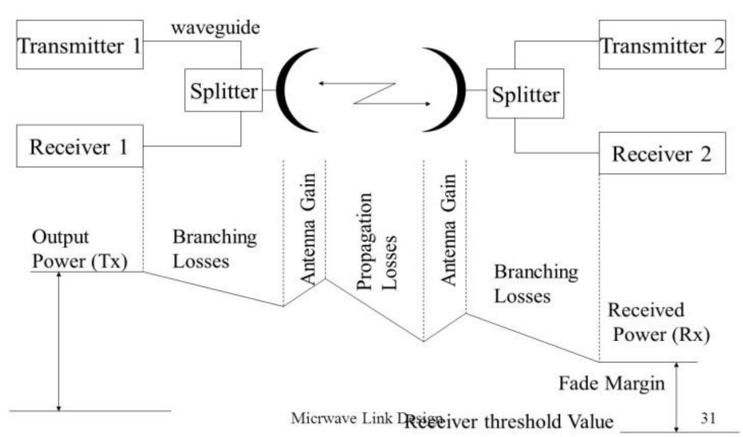


Fig 1.1: Radio Path Link Budget [1]

For a line of sight radio system, a link budget equation might look like this:

$$P_{RX} = P_{TX} + G_{TX} - L_{TX} - L_{FS} - L_{M} + G_{RX} - L_{RX}$$

Where,

P_{RX} = received power (dBm)

P_{TX} = transmitter output power (dBm)

G_{TX} = transmitter antenna gain (dBi)

L_{TX} = transmitter losses (coax, connectors.) (dB)

L_{FS} = free space loss or path loss (dB)

L_M = miscellaneous losses (fading margin, body loss, polarization mismatch, etc) (dB)
G_{RX} = receiver antenna gain (dBi),
L_{RX} = receiver losses (coax, connectors...) (dB)

II. Background: Fundamental Elements of LOS Microwave Radio Systems

Frequency

The carrier frequencies of LOS microwave links are usually above 200 MHz. For digital transmission, the frequency range between 1.8 GHz and 7 GHz is utilized. Higher frequency has two main advantages-First, it provides the large bandwidth necessary for high bit rate transmission. Second, high carrier frequencies are less susceptible to atmospheric effects by the transmission path.

The Fresnel Zone

Radio waves travel in a straight line, unless something refracts or reflects them. But the energy of radio waves is not "pencil thin." They spread out the farther they get from the radiating source — like ripples from a rock thrown into a pond. The area that the signal spreads out into is called the Fresnel zone. If there is an obstacle in the Fresnel zone, part of the radio signal will be diffracted or bent away from the straight-line path. The practical effect is that on a point-to-point radio link, this refraction will reduce the amount of RF energy reaching the receive antenna. The thickness or radius of the Fresnel zone depends on the frequency of the signal — the higher the frequency, the smaller the Fresnel zone. The reflection point offset from a direct signal path, where the length of the reflected path is exactly ½ wavelengths longer than the direct signal path. These boundaries can be calculated with the following formula:

$$F_n = \sqrt{\frac{n\lambda d_1 d_2}{d_1+d_2}}$$

Where,
F_n = The nth Fresnel Zone radius in meters.
d_1=The distance of P from one end in meters
d_2 = The distance of P from the other end in meters.

The cross section radius of the first Fresnel zone is the highest in the center of the RF LOS which can be calculated as:

$$r = 17.32\sqrt{\frac{D}{4f}}$$

Where,
r = radius (m),
D = total distance (km),
f = frequency transmitted (GHz).

Absorption

Transmitted EM energy can converts into another form e.g. thermal. The conversion takes place as a result of interaction between the incident energy and the material medium, at the molecular or atomic level. One causes of signal attenuation due to absorption by walls, precipitations (rain, snow, sand) and atmospheric gases.

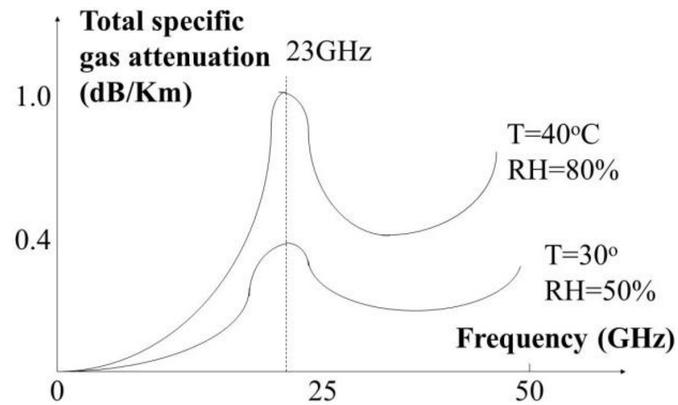


Fig 1.2: Gas Attenuation vs Frequency [Google image]

Atmospheric Refraction

The atmosphere changes dynamically and is never constant. Keep this principle in mind; we discuss the effects of atmospheric refraction, which significantly affects radio signal propagation. The result is a signal path that normally tends to follow earth curvature, but to a lesser degree. In radio engineering, atmospheric refraction is also referred to as "the K factor," which describes the type and amount of refraction. A K=1 describes a condition where there is no refraction of

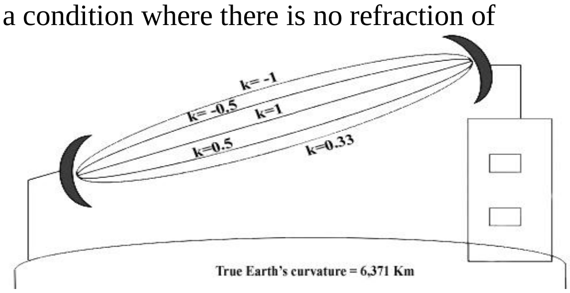


Fig 1.3: Variation of Ray curvature as a function of k [1]

the signal, and it propagates in a straight line. A K<1 describes a condition where the refracted signal path deviates from a straight line, and it arcs in the direction opposite the earth curvature. A K>1 describes a condition where the refracted signal path deviates from a straight line, and it arcs in the same direction as the earth curvature.

Effective Earth curvature or Bulge

Effective earth bulge represents the effects of atmospheric refraction, or K, combined with physical earth bulge. The following formula can be used to compute "effective earth bulge," in meter, at any data point in a path. It includes the effects of the applicable K factor:

$$h = 0.078\,\frac{d^2}{4K}$$

Where,
h = curvature height in meter,
d = path length in kilometer

System gain

This parameter incorporates all the gains and losses of the system and also determinates the transmitter power required, based on a pre-established receiver sensitivity for a given Bit Error Rate (BER) or determines receiver sensitivity based on available transmitter power at a given BER.

The system gain is given by-

$$G_{sys} = FM - G_t - G_r + L_p + L_f + L_{br}$$

Where,
G_{sys} = System Gain(dB).
FM = Fade Margin(dB).
G_t = Transmitter antenna gain(dB).
G_r = Receiver antenna gain(dB
L_p = Free space loss or path loss(dB).
L_f = Feed loss(dB).
L_{br} = Branching loss(dB).

Fade margin

Fading is define as the variation of the strength of a received radio career signal due to atmospheric changes and/or ground and water reflections in the propagation path. Fade margin is based on the link power budget. Fade margin of a link is given by-

$$FM\ (dB) = 30\log(d) + 10\log(A) + 10\log(B) + 10\log(f) - 30$$

Where,
f = career frequency (MHz).
d = path/hop length (km).
A = factor determining the terrain roughness of the path.
B = factor determining atmosphere impact on the link.

Free space path loss

Free-space path loss (FSPL) is the loss in signal strength of an electromagnetic wave that would result from a line-of-sight path through free space, with no obstacles nearby to cause reflection or diffraction. It does not include factors such as the gain of the antennas used at the transmitter and receiver, nor any loss associated with hardware imperfections. For typical radio applications, it is common to find frequency (f) measured in units of MHz and distance (d) in km, and thus FSPL equation becomes-

$$\mathrm{FSPL(dB)} = 20\log_{10}(\mathrm{d}) + 20\log_{10}(\mathrm{f}) + 32.44$$

III. Calculation

The following parameters need to calculate.

Noise power

The term noise power has the following meanings the measured total noise per bandwidth unit at the input or output of a device when the signal is not present. Noise power of a microwave link is given by-

$$N(dB) = 10\log(K) + 10\log(T) + 10\log(BW) + NF(dB)$$

Where,
N = Noise power in dB
T = Temperature in K.
BW = Bandwidth in H
NF = receiver Noise Figure in dB.
K = Boltzmann's constant.

Carrier-to-Noise ratio

The Carrier-to-Noise ratio, often written CNR or C/N, is a measure of the received carrier strength relative to the strength of the received noise. If the incoming carrier strength in microwatts is P_c and the noise level, also in microwatts, is P_n, then the carrier-to-noise ratio, C/N, in decibels is given by the formula,

$$C/N = 10\log(P_c/P_n)$$

Channel capacity

Channel capacity is the tightest upper bound on the amount of information that can be reliably transmitted over a communications channel. In microwave communication link it depends on the spectral efficiency of the selected modulation scheme and the available bandwidth of the link.

$$f_b = \text{Spectral efficiency} \times \text{Available bandwidth}$$

Modulation scheme

The choice of digital modulation scheme will significantly affect the characteristics, Performance and resulting physical realization of a communication system. In our consideration we prefer the QPSK for our link design because QPSK is widely used for its attractive power/bandwidth performance for acceptable levels of error.

Antenna tower height

For antenna height calculation in an LOS microwave link, the knowledge of obstacle heights and location along the path, earth curvature and Fresnel zone clearance are absolutely necessary.

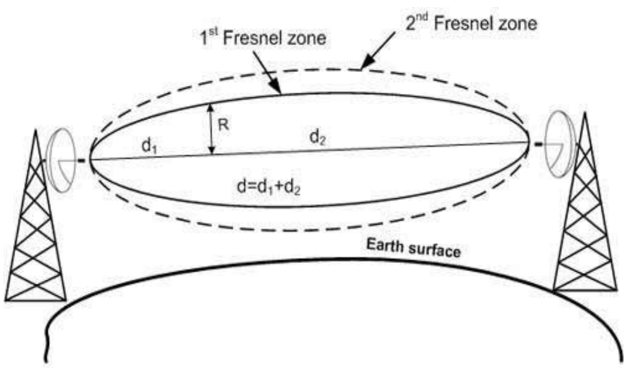


Fig 1.4: Antenna Tower Height [Google image].

The Minimum antenna tower height (H) equation is

$$H = h_m + r_m + h_{ph}.$$

Where,
h_{ph} = Physical obstacle Height (m).
h_m = Earth curvature $= 0.078\frac{d^2}{4K}$
Where,
K = k-factor value.
d = path length in km.

First Fresnel clearance,

$$r_m = 8.65\sqrt{\frac{d}{f}}$$

Where,
d = path length (km).
f = carrier frequency (GHz).

Bit Error rate (BER)

Bit error rate (BER) is the percentage of bits that have errors relative to the total number of bits received in a transmission, usually expressed as ten to a negative power. A BERT (bit error rate test or tester) is a procedure or device that measures the BER for a given transmission.

Antenna Gain

Antenna gain is the ratio of how much an antenna boosts the RF signal over a specified low-gain radiator. Antennas achieve gain simply by focusing RF energy. Gain for a parabolic antenna-

$$G = 10\log\left(\frac{4\pi f^2 A_{eff}}{c^2}\right)$$

Where,
A_{eff}= Effective Area,
f = Carrier frequency,
c = Speed of light.

Effective area, $A_{eff} = K\left(\frac{\pi D^2}{4}\right)$

Where,
K = antenna efficiency factor,
D = antenna diameter.

Transmitted Power

The transmit power is the RF power coming out of the antenna port of a transmitter. It is measured in dBm, Watts or milliWatt and does not include the signal loss of the coax cable or the gain of the antenna. Equation for the calculation is-

$$P_t = G_{sys} + 10\log(P_r)$$

Number of voice channel

The following equation is used for the calculation of the Number of voice channel.

$$N = 24\frac{f_b}{1.544}$$

Energy bit-to-noise ratio

Equation for the Energy bit-to-noise ratio is-

$$(E_b/N_0) = CNR + 10\log(BW) - 10\log(f_b)$$

Error power (P_e)

Total error power equation is-

$$P_e = e^{-CNR(\sin(450))^2}$$

The total process can be summarize in fig.1.5 as below

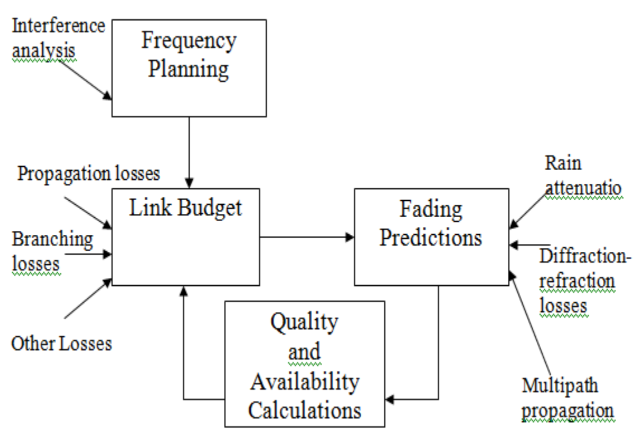


Fig 1.5: Summary of the whole process [1]

IV. Design Specification

A complete digital LOS link design.

Here we design the complete LOS link with the following link parameter and calculated the parameters with the program. The following data can be collect from survey and the link demand.

- Operating frequency, f = 2.0 GHz.
- Available bandwidth, BW = 20 MHz.
- Path length, d = 55 km.
- Antenna diameter, D = 1.5 m.
- Receiver noise figure selected, NF = 5 dB.
- Modulation scheme selected : QPSK (spectral efficiency = 1.9 b/s/Hz)
- Receiver sensitivity P_{min} = 10 pW.

Side characteristics:

- Terrain roughness: Average (A = 1.00).
- Atmospheric impact factor: Average (B = 0.25).
- Tallest physical obstacle (h_{ph}): 30m
- Environmental temperature = 300K.
- K-factor value of the terrain = 1.05.

V. Program

We developed a software program with C language to calculate the different parameters of a digital line-of-sight microwave link. GUI is given by using dotnet.

Fig-1.6: Introducing Interface of LOS software

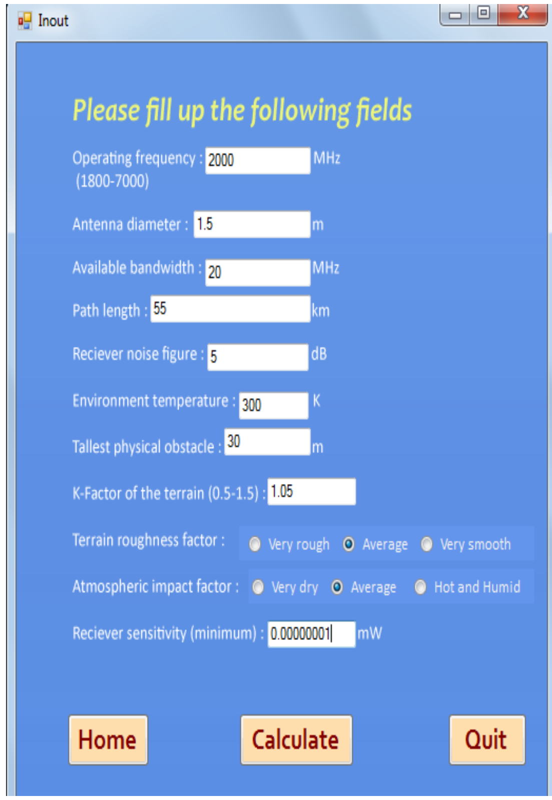


Fig-1.7: Input Interface

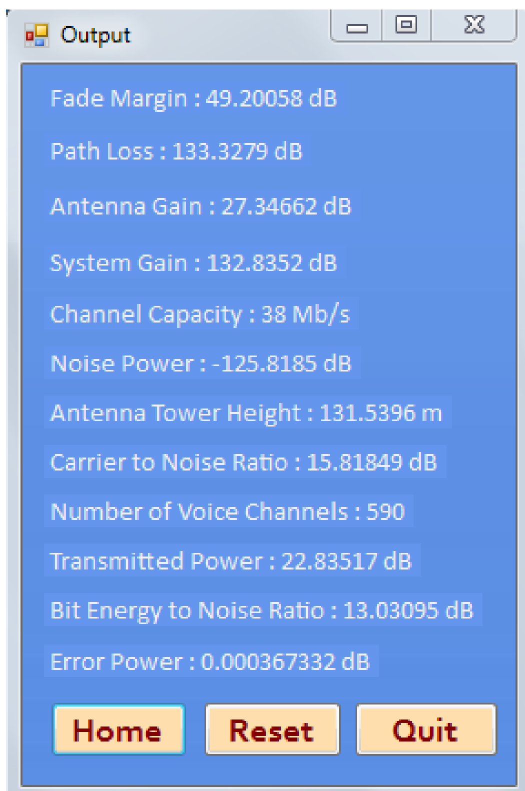


Fig- 1.8: Output Interface

INPUT:

Press Start button
Please enter the following data.
Operating frequency in MHz (2000-60000): 2000
Antenna diameter (in meter): 1.5
Available bandwidth in MHz: 20
Path length in kilometer: 55
Receiver noise figure in db: 5
Environment temperature in K: 300
Tallest physical obstacle in meter: 30
K-factor of the terrain (0.5-1.5): 1.05

Terrain	A
Very rough	0.25
Average	1.00
Very smooth	4.00

Terrain roughness factor from table (A): 1.00

Atmosphere	B
Very dry	0.125
Average	0.25
Hot and humid	0.5

Atmospheric impact factor from table (B): 0.25
Receiver sensitivity (min) in miliwatt: 0.00000001

OUTPUT:

(1) Fade margin: 49.200581 dB
(2) Path loss : 133.327850 dB
(3) Antenna gain: 27.342220 dB
(4) System gain: 132.843994 dB
(5) Channel capacity: 38.000000Mb/s
(6) Noise Power: -125.818489 dB
(7) Antenna height (minimum): 131.539551 m
(8) Carrier_to_Noise ratio: 15.818489 dB
(9) Number of voice channels: 590.

VI. Conclusion

The parameters of the link may vary on many site characteristics such as environmental temperature, reflectivity, terrain roughness, humidity, snow falling, dense foggy weather and huge rainy weather. So the calculation should be adapted for those sites depending on environmental parameters. The system gain may get changed while flying obstacles (like flying bird, aero plane etc.) come on the path of the beam though it is a transient event. We assumed here the branching loss and fading loss of a value of 2.5 dB but in practice it is a variable parameter and it depends on wave guide and antenna. Here we consider LOS between two fixed height antennas. Further study can be done considering variable tower height.

References

[1].Lehpamer, Harvey. *Microwave Transmission Networks (Planning, Design and Deployment)*.New Delhi : Tata McGraw-Hill, 2010.

[2].Kolimbiris, Harold. *Digital Communication System*. Delhi : Pearson Education, 2002.

[3].Lee, William C. Y. *Mobile Cellular Telecommunication*. New Delhi: McGraw-Hill, 2006.

[4].Rappaport, Theodore S. *Wireless Communication Principle and Practice*. New Delhi: Prentice Hall of India, 2009.

[5].Gupta, Sanjeeva. *Microwave Engineering*. Delhi: Khanna Publishers, 2006.

[6].Rao, R.S. *Microwave Engineering.* New Delhi: PHI Learning Private Limited, 2012

[7].Liao, Samuel Y. *Microwave Devices and Circuits*. New Delhi: Pearson Education, 2013

[8].Pozar, David M. *Microwave Engineering*. Crawfordsville,USA: John Wiley and Sons, INC. 2014

[9].Sisodia, M.L. et al *Microwaves-Introduction to Circuits, Device and Antennas*. New Delhi: New age Int. Publishers 2014

[10].Jordan, Edward C. et al *Electromagnetic Waves and Radiating Systems*. New Delhi: Prentice Hall Pvt. Ltd. 2011

Design of a Circular Polarization Array Antenna Using both sided MIC technology

Tasfia Tasbin, Sabrina Abedin, Piyas Chowdhury
and Md. Azad Hossain
Dept. of Electronics and Telecommunication Engineering
Chittagong University of Engineering & Technology,
CUET, Bangladesh
isha.tasfia@gmail.com, sabrinaabedinete@gmail.com,
piyas@cuet.ac.bd, azad@cuet.ac.bd

Quazi Delwar Hossain
Dept. of Electrical and Electronic Engineering
Chittagong University of Engineering & Technology,
CUET, Bangladesh

***Abstract*— In this paper, a single-feed circularly polarized array antenna is proposed. The proposed antenna is a 2x2 array antenna where each patch element is circularly polarized. The feed network has microstrip lines and slot line where microstrip-slot branch circuit is connected in parallel. The array antenna unit is realized in very simple structure as the feeding circuit of the design does not require any external network. "Both-Sided MIC Technology" is used to design the feed network. The design frequency of the proposed array antenna is 3.8 GHz. The axial ratio below 3dB indicates that the proper circular polarization is achieved. The design and the basic operation along with the simulation results of the proposed circularly polarized array antenna are demonstrated in this paper. In the addition, The simulated S-parameters and axial ratio of the feed circuit are explained.**

Keywords—circular polarization; microstrip antenna; array antenna; Both- Sided MIC technology.

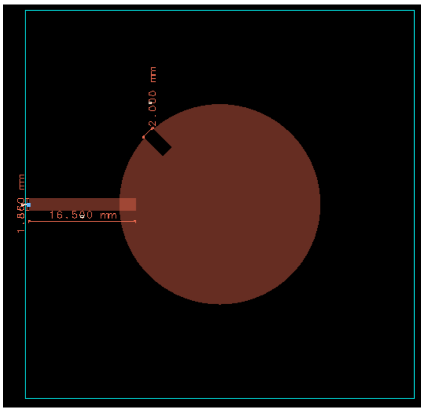
I. INTRODUCTION

Microstrip antennas (MSA) are used in a large number of microwave communication applications due to the amazing features of light weight, low volume, low cost, easy fabrication, conformal configuration and compatibility with integrated circuits [1]. In many applications, the requirement of circularly polarized MSA with a wide field of views becomes more demanding in mobile stations, satellite communications, high-resolution radar systems, and so on [19]. Among the microstrip antennas, circular polarization microstrip antennas have inherent capabilities of reduced multi-path fading, improved coverage and fixed polarization. As a result, antenna systems that employ the circular polarization are attractive to the users as they give the opportunity to enhance the Polarization. It is suitable for wireless local area network (WLAN) to mitigate the detrimental fading loss in a multipath environment [2]. Circular polarization antenna is of huge benefit for radio-frequency identification (RFID) antennas where the antennas are randomly oriented [3, 4, 5]. This also provides the facility of frequency reuse for doubling the system capability that is useful for satellite communication system [6]. Several antenna designs, offering circular polarization have been reported [13-15]. Some scheme of getting polarization diversity is by changing the electric characteristics of perturbation segments through PIN diodes have been introduced in [7-10]. The novel circular polarization detection patch array antenna is also proposed [15]. Microstrip array antenna to radiate dual circular polarization has been presented in [12]. A 2×2 slot-ring array antenna with orthogonal feed circuit is used to excite the orthogonal circular polarization in [12] but it is multilayer configuration. RHCP and LHCP can be achieved simultaneously by interchanging the ports of the feed circuit. 4-element circular polarization switchable patch array antenna is proposed here [14].

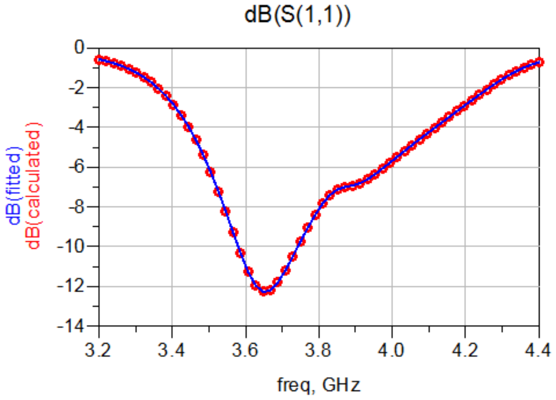
In this paper a dual circularly polarized microstrip array antenna is proposed. The proposed array antenna consists of a 2×2 circularly polarized patch antenna and orthogonal feed circuit which is designed using both microstrip lines and slot lines. The circular patch element with single slit arena has been introduced. To realize feed circuit effectively, Both-sided MIC Technology [15-18] is used. This design is also suitable for array size extension such as 4×4, 8×8 array.

II. STRUCTURE OF ANTENNA

A. Single Patch Element



(a)



(b)

Fig. 1.(a) Geometry of a single patch element and (b) performance of S(1,1)parameter for single patch element

The geometry of a single patch element is proposed and shown in Fig. 1.(a). The circular patch element with slit arena is inherently circularly polarized. Here, the circle has the radius of 15.5mm. The slit arena has a depth of 3.996mm and width of 2mm.

The required data for S (1, 1) parameter is presented in Fig.1.(b). The data of S(1,1) parameter which is below 12-dB indicates good performance of the proposed single patch element. Where, the frequency of single patch is gained around 3.65GHz. The same patch element has been used in as the four elements

B. The proposed Antenna Design:

The complete layout of the proposed circular polarization array antenna is shown in Fig. 2. The proposed antenna has four radiating patch elements which consists of every single elements which is shown in previous in Fig.1.(a) and microstrip lines, slot line. In substrate layer, Teflon with a thickness of 0.8 mm and relative dielectric constant of 2.15 is used to design the proposed antenna. The impedance of the microstrip line in the design is 50 ohm (2.4 mm).

Due to the existence of truncated corner, the size of radiating patch elements can be reduced considerably. In the layout, width of microstrip line and slot line are 2.4mm and 0.2mm respectively. The impedance of microstrip line and slot line in the layout are 50.5 ohm and 155 ohm respectively. For proper matching between microstrip line and slot line, 88.5 ohm quarter-wavelength transmission line is used as impedance transformer which is evaluated from the equation:

$$Z_0' = \sqrt{(Z_0 Z_L)}$$

Where Z_0 and Z_L is the characteristic impedance and load impedance respectively [11].

The feeding part consists of microstrip-slot branch circuit which is connected in parallel. Thus, the feeding structure has the effect of multiple feed but without complicated feeding network.

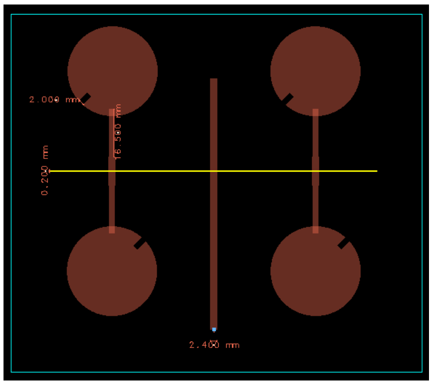
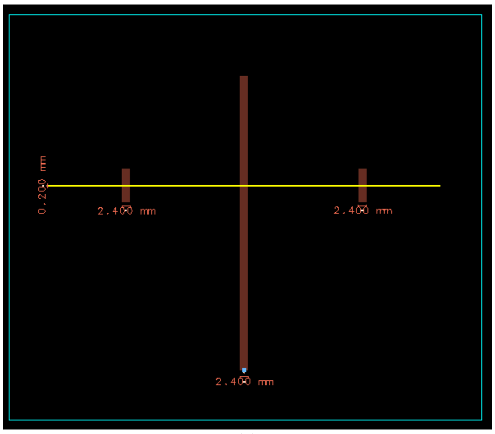


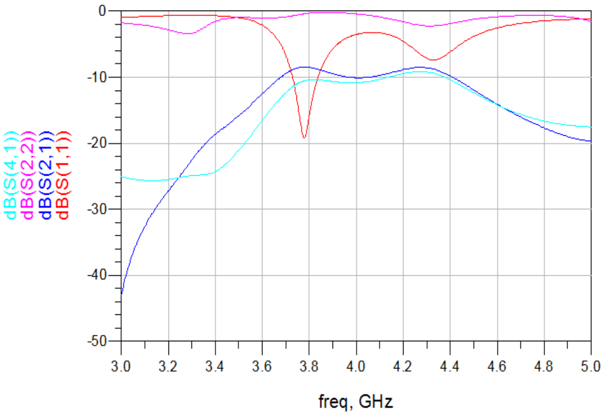
Fig. 2. Layout of the proposed CP array antenna

Orthogonal feed network is used to realize the circular polarization which is shown in Fig. 3 along with simulated results. The main two conditions for the circular polarization are two orthogonal signals and 90º phase shift between the orthogonal signals. To design this type of feed network Both - sided MIC Technology is very useful.

In the simulated results of S(1,1), S(2,1), S(2,2), S(4,1) the data which are evaluated for the feed network near to the optimized value. The data of S(1,1) parameter which is below 19-dB indicates the excellent performance of the proposed antenna. The data of other parameters also indicate good performance of the antenna. The measured antenna gain is about 6.24 dBi.



(a)



(b)

Fig. 3. (a) Layout of orthogonal feed network and (b) simulated results

C. Basic operation of the Array Antenna:

The basic operation of the proposed antenna is demonstrated:

In the patch elements, if the width of the truncated corner increases, the impedance on the Smith chart will decreases.

For circular polarization, as the patch elements are etched, this produces two resonance mode which are spatially orthogonal. Between the resonance frequencies of this two modes, a frequency of the circular polarization is obtained. When RF signals are applied to the port 1, through the microstrip-slot parallel coupled branch circuit the signals are divided equally. Then the signals are divided in anti-phase to the microstrip lines in the radiating patch elements. When the signals are fed to the each patch element then the circular polarization occurs. As one port is connected in this design, so only left-hand circular polarization (LHCP) occurs. If another port will be connected, right-hand circular polarization (RHCP) occurs [13].

Some related studies use two individual feed networks in different layers to excite separate radiating elements [6]. These designs require a multilayer structure for implementing dual-frequency arrays and are very well suited for dual-frequency operation with a large frequency ratio. But dual frequency array antenna requires relatively more complicated feed network when good impedance matching at two separate operating frequencies is required where the proposed antenna requires simple structure.

D. Simulation results:

The simulated results are shown in the following figure.

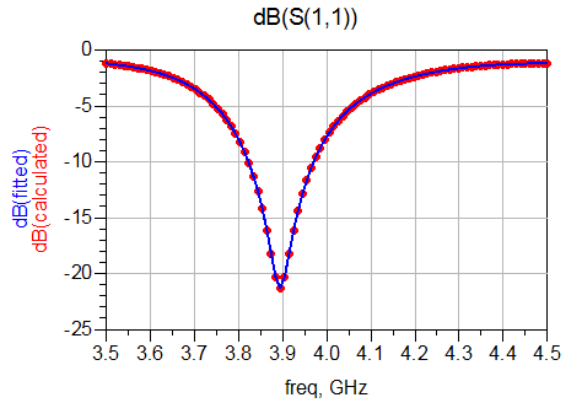


Fig. 4. Data of S(1,1) parameter

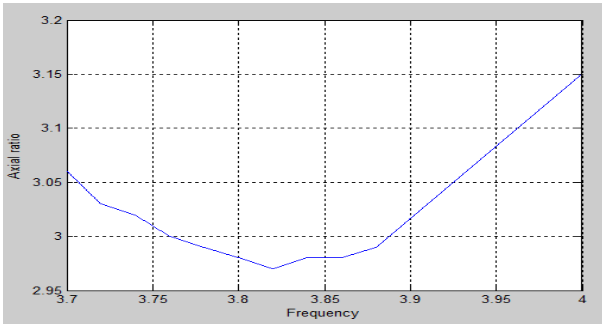


Fig. 5. Data of Axial ratio

The required data for S (1, 1) parameter and the Axial ratio is presented in Fig. 4 & 5 respectively. Data of axial ratio which is illustrated in Fig. 5. This result shows a good axial ratio below 3-dB at the design frequency which indicates the proper circular polarization is achieved. The circular polarization occurs between the frequency of 3.76 to 3.88 GHz. The data of S(1,1) parameter which is below 20-dB indicates good performance of the proposed antenna. At resonant frequency the efficiency is determined that 99.67% and the measured antenna gain is about 6.24 dBi.

III. CONCLUSION:

In this paper, a circular polarization 2x2 array antenna is proposed and the characteristics of the proposed antenna are determined. The major advantage of single-feed, circularly polarized microstrip antennas is their simple structure, which does not require an external polarizer. Therefore, they can be realized more compactly by using less board space than dual-feed circularly polarized microstrip antennas. Each radiating patch element of the proposed antenna is circularly polarized and LHCP is produced using the proposed antenna. Axial ratio below 3-dB indicates that the proper circular polarization occurs. All the simulated results indicate that the proposed array antenna radiation is circularly polarized with good return loss performance. The most tremendous feature of this array antenna is that it can be implemented to array size 4x4, 8x8 etc. Recently, in mobile base stations, wide-angle coverage satellite communications, high-resolution radar systems also requires dual circular polarization where polarization diversity is needed. This proposed array antenna is perfectly suitable for practical wireless communication and communication sectors.

REFERENCES

[1] Hossain Md. Azad, Eisuke Nishiyama, and Masayoshi Aikawa "Gain Enhanced Linear Polarization Switchable Microstrip Array Antenna" (978-1-4244-4968-2/10/2010 IEEE)

[2] S.-T. Fang, "A novel polarization diversity antenna for WLAN application," (2000 IEEE AP-S Dig., July 2000, pp. 282–285.)

[3] M. A. Kossel, R. Kung, H. Benedickter, and W. Biichtokd, "An active tagging system using circular-polarization modulation," Trans Microw.Theory Tech., vol. 47, no. 12, pp. 2242–2248, December 1999.

[4] M. Kossel, H. Benedickter, and W. Baechtold, "Circular polarizedaperture coupled patch antennas for an RFID system in the 2.4 GHz ISMband," IEEE Radio and Wireless Conference (RAWCON 99), pp. 235-239, Aug. 1999.

[5] J. Garcia, A. Arriola, F. Casado, X. Chen, J. I. Sancho, and D. Valderas, "Coverage and read range comparison of linearly and circularly polarized radio frequency identification ultra-high frequency tag antennas," IET Microwaves, Antennas and Propagation, vol. 6, no. 9, pp.1070–1078, 2012.

[6] Kin-Lu Wong Copyright_c 2002 John Wiley & Sons, Inc. *"*Compact and Broadband Microstrip Antennas.*"* (ISBNs: 0-471-41717-3 (Hardback); 0-471 22111-2 (Electronic)) (chapter:5)

[7] R.-H. Chen and J.-S. Row, "Single-Fed Microstrip Patch Antenna With Switchable Polarization," IEEE Trans. Antennas Propagat., vol. 56, no. 4, pp. 922-926, Apr. 2008.

[8] F. Yang and Y. Rahmat-Samii, "A reconfigurable patch antenna using switchable slots for circular polarization diversity," IEEE Microw. And Wireless Compon. Lett., vol. 12, no. 3, pp. 96–98, Mar. 2002.

[9] M. K. Fries, M. Gräni, and R. Vahldieck, "A Reconfigurable Slot Antenna With Switchable Polarization," IEEE Microw. and Wireless Compon. Lett., vol. 13, no. 11, pp. 490-492, Nov. 2003.

[10] M. Boti, L. Dussopt, and J.-M. Laheurte, "Circularly polarized antenna with switchable polarization sense," Electron. Lett., vol. 36, no. 18, pp. 1518–19, August 2000.

[11] Elements of Electromagnetics by Matthew N.O. Sadiku. (chapter:11;article:11.6)

[12] Y. Ushijima, S. Feng, E. Nishiyama and M. Aikawa, "A Novel Circular Polarization Switchable Slot-Ring Array Antenna with Orthogonal Feed Circuit," Asia-Pacific Microwave Conference, APMC, 2010, pp. 1569-1572.

[13] Muhammad Asad Rahman, Quazi Delwar Hossain, Md. Azad Hossain, Md. Moshiul Haque Eisuke Nishiyama, Ichihiko Toyoda "Design of a Dual Circular Polarization Microstrip Patch Array Antenna" (978-1-4799-6062-0/14/$31.00©2014 IEEE)

[14] Md. Azad Hossain, Piyas Chowdhury, Quazi Delwar Hossain, Eisuke Nishiyama and Ichihiko Toyoda " Design of a Circular Polarization Switchable Microstrip Array Antenna using Magic-T Bias Circuit."

[15] M. A. Hossain, Y. Ushijima, E. Nishiyama, I. Toyoda, and M. Aikawa "Orthogonal Circular Polarization Detection Patch Array Antenna using Double-Balanced RF Multiplier" (Progress In Electromagnetics Research C, Vol. 30, 65{80, 2012*)*

[16] Kodama, K., E. Nishiyama, and M. Aikawa, "Slot array antenna using both-sided MIC technology," *IEEE International Symposium on Antenna and Propagation*, Vol. 03, 2715-2718,2004.

[17] M. A. Hossain, E. Nishiyama, M. Aikawa, and I. Toyoda "Multi-Band Orthogonal Linear Polarization Discrimination Planar Array Antenna" (Progress In Electromagnetics Research C, Vol. 34, 2013.)

[18] Egashira, K., E. Nishiyama, and M. Aikawa, "Microstrip array antenna using both-sided MIC feed circuits," (Asia-Pacific Microwave conference, APMC, 2002.)

[19] Xiong Ye, Mang He, apingyuan Zhou and Houjun Sun "A Compact Single-Feed Circularly Polarized Microstrip Antenna with Symmetric and Wide-Beamwidth Radiation Pattern" (International Journal of Antennas and PropagationVolume 2013 (2013), Article ID 106516, 7 pages)

Smart Window Blind Control System

M. Bhuiyan[1], D. A. Firdhaus[2] and Kh. Arif Shahriar[3]
[1]*Department of Electrical and Electronic Engineering, Northern University Bangladesh, Dhaka, Bangladesh.*
[2]*Department of Mechatronics Engineering, International Islamic University Malaysia, Kuala Lumpur, Malaysia.*
[3]*Department of Electrical and Electronic Engineering, Northern University Bangladesh, Dhaka, Bangladesh.*
[1]bmoinul1972@yahoo.co.jp, [2]dzulasyraff@gmail.com and [3]arif.sjn@gmail.com

***Abstract*—This paper presents a smart window blind system to control correct amount of sunlight needed to shine a room. The system can be operated in either manual or automatic mode. The manual mode activates two push buttons for the tilting control. For automatic mode, the blind will open or close, depending on the amount of light that is shining through the window. The blind will be closed fully when the sensor level is below 200, and will be opened at 50% when the value is above 800. If the sensor level is between 200 and 800, the blind level will be fully opened. A servo motor is attached to the blind's controlling shaft for the actuation of tilting. The results show that the system is usable and can be used commercially in office and home.**

Keywords*— *Venetian blinds, glare, ambient light sensor, IR remote controller, servo motor.

I. INTRODUCTION

Optimum utilization of daylight, maintaining proper indoor lighting intensity and glare protection are some of the challenges that need to be solved to ensure a comfortable ambience at homes and offices. A window blind gives a high level of user convenience in terms of heat protection and also prevents glare related problems. Direct emission of sunlight on the furniture will eventually damage them. Automation of the window blind allows the user to protect the valuable furniture from direct sunlight. The controlling system that regulates the blind opening and closing during day time can either be computer based or microcontroller based [1] [2].

Apart from that, when the sun goes down, the windows become transparent and the private business becomes visible to anyone passing by. The ability to automatically close the blinds after sunset can be achieved by using ambient light sensor.

In addition, the automation of window blind would save energy consumption and will ensure the sustainability of the environment [3]. Closing the blind automatically during the hot weather will prevent the hot sun rays from heating up the room and thus will reduce the usage of air conditioners. Likewise, when it is cold outside, closing the blind can help to trap the valuable heat inside. If automated and controlled correctly, solar blind systems can reduce energy requirements by 10 percent.

The blind systems are used to filter completely or partially the daylight entering into the room. To prevent glare sources from impacting the occupancy area, some constraints on working state of the shading device are defined. The working state of a venetian blind is described by the descent of the blind and the angle of the slats. To counteract a glaring ray, a minimum descent (denoted by d_{min}) and a minimum angle of slats (denoted by θ_{min}) are imposed to the blind system [4]. It is as shown in Fig.1.

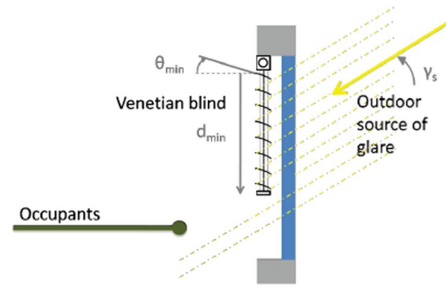


Fig.1: Venetian blind position for glare rejection.

The health and safety of user should never be compromised in order to reduce costs. The blind should ensure the privacy of the user staying inside the room. For example, during night, the blinds must be closed and opaque so that people from outside are unable see what is happening inside the room. The smart window blind should be built to have the longest possible useful life and require minimum repairing and maintenance. In case of damage, the replacement parts should be easily available and the repairing process should be easy for the user.

The position of the sun will affect the smart window design. At early morning and late afternoon, the light level is usually ambient and hence the blind should totally open. At mid-day, the sunlight will usually be intense enough to cause glare to the user and affect his/her vision. Then, the blinds should tilt down and allow small amount of light into the room. The blind should fully close at night in order to protect the privacy of the user.

A lot of researches have been conducted on window blind control. Chen *et al.* presented the venetian blind control using fuzzy neural network for indoor day lighting [5]. Solar powered smart blind system was proposed by Herrera *et al.* [6]. Kim *et al.* have conducted an extensive experimental study to evaluate the environmental performance of automated venetian blind [7]. As the Microcontroller Unit (MCU) is currently popular in designing various automatic and autonomous systems [8] [9] [10], this research uses the Arduino microcontroller unit for designing the motion control system of the window blind.

II. SYSTEM ANALYSIS

The smart window blind uses the concept of wireless communication which is a method of controlling the tilting of the blind with a motor without direct wiring. The motor connected to the controlling shaft and the gear provides balance and gripping to the shaft that controls the blind's tilting angle. Fig.2 shows the flow-chart of the algorithm of the smart window blind system.

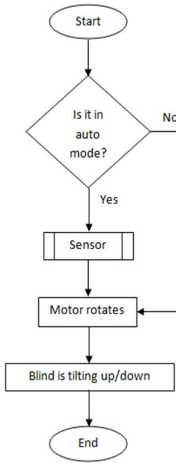


Fig.2: Flow-chart of the algorithm for operation of the smart window blind system

The user can choose one from two modes - either automatic or manual. The automatic mode uses sensor attached at the front of the window. The sensor will detect the intensity of light and automatically tilts the blinds to an appropriate angle. On the other hand, the manual mode provides a direct control from the user via a handheld remote controller. The user will tilt the blind based on his/her own preference. This mode is preferred whenever automatic mode can't provide optimum amount of lighting into the room.

The code is written in C programming language. The algorithm is comprised of a process that monitors the digital values of the controls on the switch box. The system monitors a push button for the mode selection - either automatic or manual. The rocker switch is used as the primary input to the system. If the switch sends out binary 0, the system will be activated in automatic mode while sending out binary 1 will cause the system to operate in manual mode. The mode information will be displayed on LCD to let the user know the current state of the blind.

In automatic mode, the optical sensor returns an analog value ranging from 0 to 1023. The value 0 indicates the darkest scenario while 1023 indicates the brightest. Three ranges of analog values which represent the ambient, bright and dark environment will be declared in the microcontroller.

Every second, a new value is read from the optical sensor and based on the range to which value belongs, the servo motor is rotated in the appropriated direction and revolution that has been set in the program.

On the other hand, in manual mode, some sets of binary data will be declared to represent specific commands. Then, the signal send out by infrared transmitter will be detected by the receiver that has been connected to the microcontroller. The signal received will be decoded by the microcontroller into a binary data. These binary data will be compared to the default data in the programming and the actuator will carry out the command.

III. DESING AND CALIBRATION

Smart window blind mainly uses hardware for the design of the controller. The system also uses a rocker switch for user selection of desired control mode. Two types of controller are used which are infrared handheld remote control and ambient light sensor.

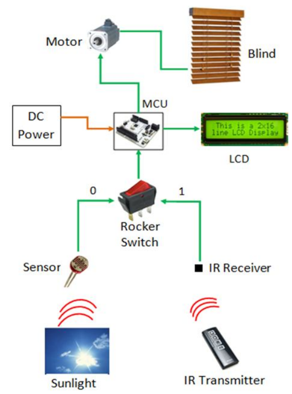


Fig.3: Hardware model of the Smart Window Blind System

Basically these two devices act as receiver. They will receive signal from the sources or transmitter. This signal will then be sent to the microcontroller which will send out the command to actuate the motor. The blind's tilting will be proportional to the total rotation made by the motor. Connections between the hardware components have been illustrated in Fig.3.

A. Smart Window Blind Design

A servo motor will be attached directly to the controlling shaft. The angles of tilting were determined by the rotational number of the motor. The microcontroller unit, infrared receiver and the sensor will be placed in a controller unit.

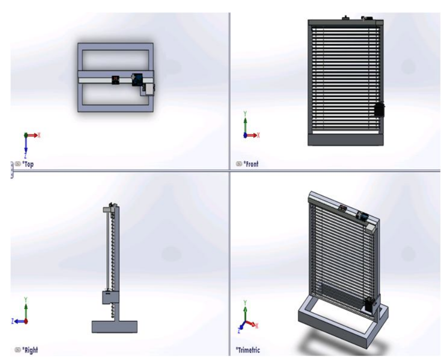


Fig.4: Smart window blind design sketched in Solid Work

The system is powered by 6V power supply. LCD would let the user know the current mode of the system - automatic or manual. A rocker switch is used for the selection of desired operating mode of the blinds. The tilting angle of the blinds will correspond to the rotational number of the servo. Fig.5 shows the whole connectivity among the different components of the control system.

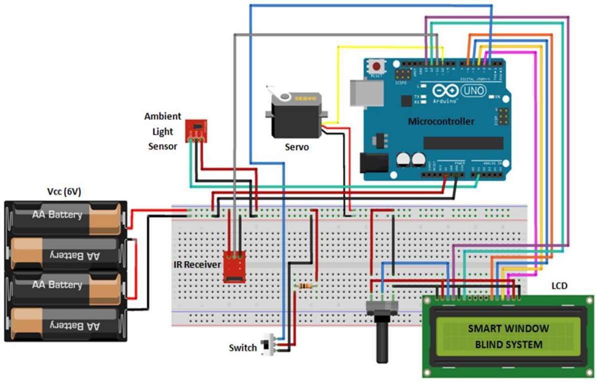


Fig.5: Designed control system circuit sketched by Fritzing

B. Smart Window Blind Design

Fig.6 shows the configuration for the calibration of the ambient light sensor. This testing is to determine whether the sensor can control the activation of the LED based on the amount of light falling on the sensor. A coding structure has been sketched via Arduino software which gives instructions to the LED to turn on whenever the light level is above 200 and off for light level below 200. Finally, a light source is used to vary the amount of light falling on the sensor.

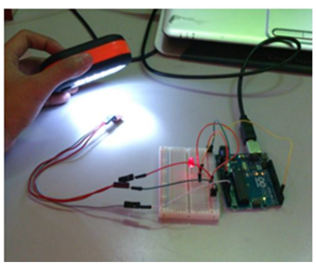


Fig.6: Configuration for the calibration of sensor

The test shows that the sensor can control the activation of LED as expected. When the sensor is exposed to the dark room, the LED is off. LED is on whenever the lamp is directed by the sensor. The calibration of IR remote controller has been done by sketching a set of coding into Arduino which gives instruction to LED to turn on whenever button 'A' is pushed and turn off when button 'B' is pushed. Fig.7 shows the configuration for the calibration of IR remote controller.

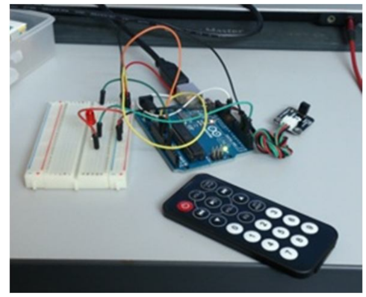


Fig.7: Configuration for the calibration of IR remote controller kit

C. Calibration of Ambient Light Sensor

The analog values of sunlight in a day have to be known so that the ranges of light level can be designed. The light sensor has been set up and the analog values of light in a day has been recorded.

Table 1: Analog Readings of Light Intensity by Sensor on 1/5/2013

Light Levels over Time (on 1/5/2013)												
Time (am)	12:00	1:00	2:00	3:00	4:00	5:00	6:00	7:00	8:00	9:00	10:00	11:00
Analog Value	7	8	8	11	24	52	74	198	306	561	708	918
Time (pm)	12:00	1:00	2:00	3:00	4:00	5:00	6:00	7:00	8:00	9:00	10:00	11:00
Analog Value	1018	1023	1021	983	872	551	409	87	35	22	12	9

The light sensor has been exposed to sunlight for a whole day and the analog reading for every hour was recorded. Value 0 indicates that the brightness level is at lowest while 1023 indicates the brightness level is at the highest.

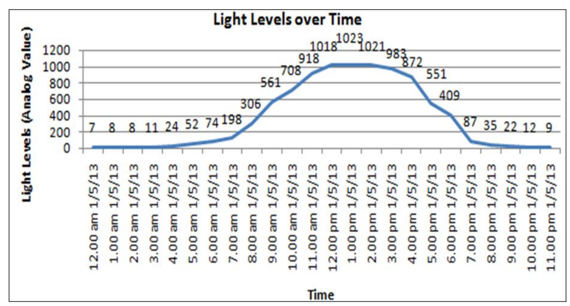


Fig.6: Graph of light levels over time

Based on the calibration, the total system settings has been declared and integrated to the main system. Table 2 shows the configuration for automatic mode.

Table 2: Automatic mode system configuration

Input (Analog Values)	Light Level	Output (Tilting)
0 ~ 200	Dark	Fully close
201 ~ 800	Ambient	Fully open
801 ~ 1023	Bright	45° opening

Normally, the sun rises at 7.00 am and the brightness level is ambient. Thus, the blind should fully open and allow the light into the room. When the time is 11.00 am, the sun is positioned almost at the peak of its trajectory. Then, the surrounding will be brightest and glaring effect will rise. At this moment, the blinds should allow only some amount of light into the room by opening at 45 degree. Later, at night, there is no sunlight and the outside is dark. The inside part of the room will be transparent to the outsider. For privacy concern, the blind will be fully closed.

D. Calibration of IR Remote Controller

Similar to ambient light sensor, each button of the IR remote controller was represented by analog values. These analog values are converted into decimal and hexadecimal. A simple configuration has been set up to extract the analog values of the buttons.

Table 3: Codes of IR remote controller

Decimal	Hexadecimal	Button (Symbol)
16584943	FD10EF	V
16601263	FD50AF	^
16617583	FD906F	VOL -
16625743	FDB04F	EQ
16609423	FD708F	ST/REPT
16593103	FD30CF	0
16582903	FD08F7	1
16615543	FD8877	2
16599223	FD48B7	3
16591063	FD28D7	4
16623703	FDA857	5
16607383	FD6897	6
16586983	FD18E7	7
16619623	FD9867	8
16603303	FD58A7	9

Table 4 shows the configuration for manual mode which is integrated to the main system.

Table 4: Manual mode system configuration

Input (Button/Symbol)	Motor Rotation	Output (Tilting)
Up (∧)	Counterclockwise	Open
Down (∨)	Clockwise	Close

IV. RESULT AND ANALYSIS

A. Development of Smart Window Blind Prototype

Refer to Fig.7 (a); the black frame represents the room's window and the orange box represents the main processing unit of the system. The controlling unit has been put at the lower front part of the blind so that the user can easily communicate with the system.

Most of the electrical components were placed around the controlling unit box as shown in Fig.7 (c) so that it is easier to debug the system. The connection among the components also has been compiled at the inner part of the controlling unit because to avoid high temperature, high humidity and other disturbance that may affect the system performance. Both input and output units were connected to the controlling unit. An LCD has been attached to the controlling unit as to inform the user about the current operating mode of the system; either automatic or manual.

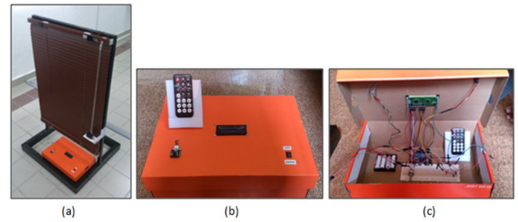


Fig.7: Smart window blind prototype: (a) Final prototype, (b) Controller unit, (c) Inside of controller unit

Fig.7 (b) shows the input units of the smart window blind system. A rocker switch has been fitted to the front part of the controlling unit so that the user can easily achieve and choose an operating mode. The rocker switch has markings of "auto" and "manual" for convenience of the user. An IR remote and receiver have also been placed at the front part of the system to make the interaction between the remote and the receiver smooth without any barriers.

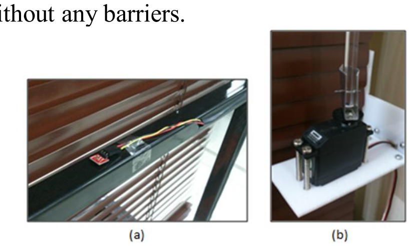


Fig.8: Outputs of smart window blind system (a) Ambient light sensor on the frame, at the back of window blind, (b) Servo directly attached to tilt-controlling shaft

As for auto mode, the sensor has been fitted at the back part of the system as shown in Fig.8(a), which also symbolizes that the sensor would face up to the outer part of the room and the sunlight. The light falling on the window will be detected and converted by the sensor as the input of the system. Servo motor is attached directly to the controlling shaft for the actuation of tilting as shown in Fig.8 (b).

B. Testing and Observation

Lab test is conducted, shown in Fig.9, as to observe whether the blinds can cooperate with the three states when it is operating in automatic mode. At first, all sources of lights in the lab have been turned off. As a result, the servo rotated and the blind was fully closed. This configuration is to imitate darkness. Next, the light bulb of the lab has been switched on as to imitate the environment with ambient lighting. Then, the blind was fully opened by itself. Lastly, a lab lamp has been used as a source for the direct lighting to the window. Since the light intensity was too high, the blind then changes to 45 degree opening as predicted.

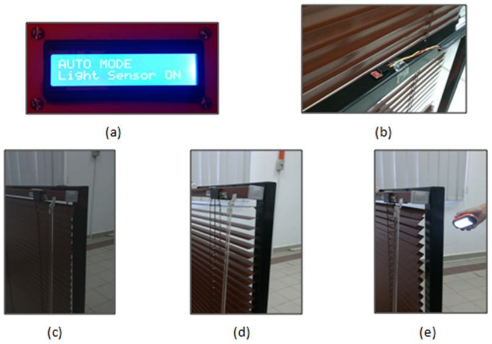


Fig.9: Automatic mode lab test: (a) Auto mode activated, (b) Light sensor on the window, (c) Blind closed when dark, (d) Blind fully opened when ambient, and (e) Blind opened to 45 degree when bright

Manual mode also needs to be tested as shown in Fig.10. In this mode, the blind will open up and close down based on the user preference regardless of the outside brightness level. The testing was carried out by pressing any one of the two buttons that have been set up for controlling of the system. Button 'A' will rotate the motor counterclockwise and open up the blind while button 'B' will rotate the motor clockwise and close down the blind.

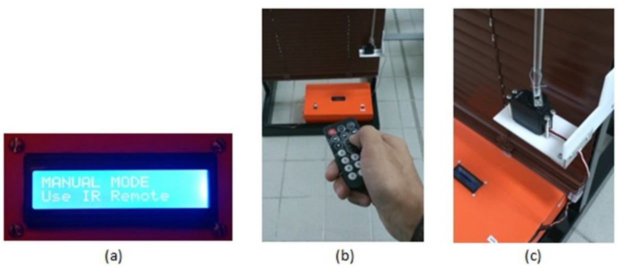


Fig.10: Manual mode lab test: (a) Manual mode activated, (b) Interaction between IR remote unit and receiver, and (c) Servo rotates based on which button was pushed

In addition, the selection of operation in automatic and manual mode was determined by the activation of the rocker switch. The LCD was displaying the information about the current state of the system; automatic or manual mode.

C. Real Time Behavior

Malaysia is a country that has equatorial climate with high temperatures and wet months throughout the years. A testing has been carried out by exposing the blind to the sunlight. For a whole day, the light level detected by the sensor for each hour is recorded and the opening of the blind is observed. The testing has been done on both dry and rainy days. Then, the effectiveness of the blind adjusting the opening by itself in the real situation can be confirmed.

The first test was done on a dry and high temperature day. The blind's opening and closing durations were observed. The data have been summarized in Fig.11.

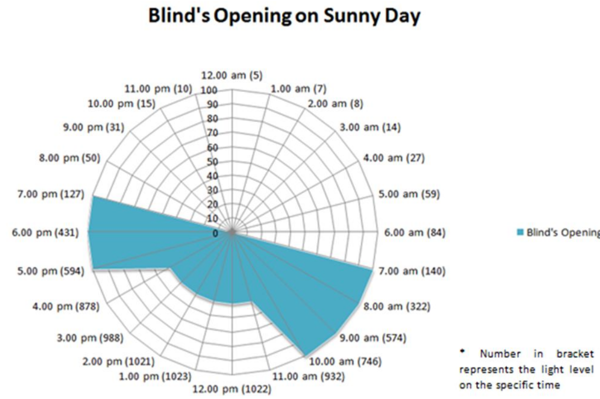


Fig.11: Blind's opening on sunny day

From 7 pm to 7 am, the brightness level is very low and people from outside can see what's happening inside the room. Thus, the blind was fully closed for the security purposed. Then, after 7 am, it is sunrise, the outside was bright and the blind starts to open up. The opening was at maximum to allow all the light into the room. However, at 10 am the sun is positioned at peak, high amount of light falls on the window and produces glare effect to the user inside of the room. The blind only opened to half of the maximum opening to allow appropriate amount of light into the room. Finally, at 7 pm as the outside was getting dark, the blind becomes fully closed.

Then, the second test was done to observe the opening of the blind on the wet, rainy day. Basically, on that day, the weather was rainy and the outdoor brightness levels detected by the sensor were either at medium or low. Fig.12 shows the light level throughout the rainy day and the effect on the blind's opening. The blind was either fully opened or fully closed since the day was gloomy and only small and moderate amount of light can be detected by the sensor.

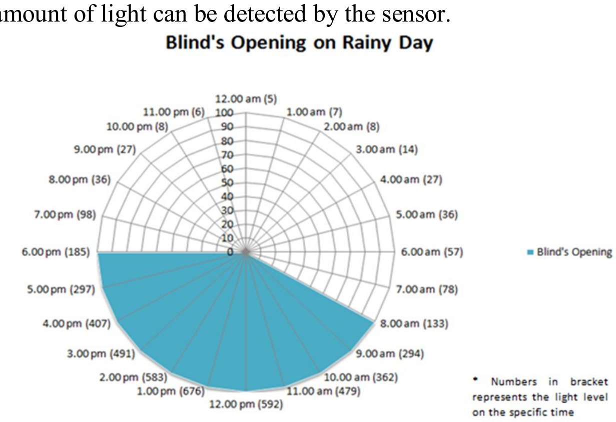


Fig.12: Blind's opening on rainy day

From these real time tests, the smart window blind system has been proved to work properly and can adapt to the real situation. The system could tilt the blind based on the amount of light falling on the window. During the dry season, the blind could reject the glares produced by the sunlight on the brightest day by tilting to half of the maximum opening and provide sufficient lighting to the room while during the rainy season, the blind provides a good privacy as it gets dark outside and most people are at their home.

V. CONCLUSION AND RECOMMENDATION

A smart window blind system has been developed and designed so that it can automatically adjust to the amount of sunlight shining through a window. For the automatic mode, the blinds adjusted the tilting angle of the slat based on the light intensity that has been programmed into the microcontroller. When ambient lighting falls on the window, the servo would rotate counterclockwise to fully open the blinds while when there is very low amount of light, the servo would rotate clockwise to fully close the blinds.

The smart window blind system has been designed based on the reviewing the basic concept and mechanism of common window blind. The components that are used for the construction of the prototype are analyzed for selection of the best materials and specifications that are suitable for the load and optimizing the budget. Originally, the servo should be integrated into the rail of the blind so that the shaft would not have to be coupled together as shown in the design. When there is too much ambient light and not enough direct light to accurately detect by the sensor, inaccuracies would be produced and as a result, the microcontroller would interpret it wrongly.

In addition, for future improvements, the servo should be directly integrated into the main frame of the blind and the use of the blind's shaft can be eliminated. Electrical energy also can be substituted with a renewable, clean energy such as replacing the use of batteries with solar energy.

REFERENCES

[1] Aida, J. (2010). Smart Window Shade. Bachelor Degree Project 2. Universiti Teknikal Malaysia Melaka. Melaka, Malaysia.

[2] M. Akhtaruzzaman, and A. A. Shafie (2011). A novel gait for Toddler biped and its control using PIC 16F877A. 2011 4th International Conference on Mechatronics (ICOM), 2011, Kuala Lumpur, Malaysia. pp. 1-6.

[3] M. H. Masud, M. Akhtaruzzaman, S. M. S. Bari, F. Anwar (2011). Engineers' Obligations towards Sustainable Environment. 2nd International Conference on Professional Ethics and Education in Engineering (ICEPEE), 2011, Kuala Lumpur, Malaysia, pp. 85-92.

[4] Bideaux, E., Arnal, E., & Anthierens, C. (2011). Consideration of Glare from Daylight in the Control of the Luminous Atmosphere in Buildings. 2011 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM2011). Budapest, Hungary.

[5] Chen, Y., Li, H., & Chen, X. (2009). Venetian Blind Control System Based on Fuzzy Neural Network for Indoor Daylighting. 2009 Second International Conference on Computer and Electrical Engineering.

[6] Herrera, M., Li, L., & Long Vo, K. (2009). Solar Powered Smart Blind System. ECE4007 Senior Design Project. School of Electrical and Computer Engineering. California.

[7] Kim, J., Yang, K., Park, Y., Lee, K., Yeo, M., & Kim, K. (2007). An Experimental Study for the Evaluation of the Environmental Performance by the Application of the Automated Venetian Blind. Proceedings of Clima 2007 Well Being Indoors. Seoul, Korea.

[8] M. Akhtaruzzaman, S. K. Hasan, and A. A. Shafie. Design and development of an intelligent autonomous mobile robot for a soccer game competition. Proc. of Int. Con. On Mechanical and Electronics Engineering (ICMEE 2009, World Scientific), 2009. pp. 1-5..

[9] M. Akhtaruzzaman, N. B. Samsuddin, N. B. Umar, M. Rahman. Design and development of a wall climbing Robot and its control system. 12th International Conference on Computers and Information Technology, 2009. ICCIT'09. 2009, pp. 309-313.

[10] Hossain, M., Rashid, M., Bhuiyan, M., Ahmed, S., Akhtaruzzaman, M. (2013). A qualitative approach to mobile robot navigation using RFID. In: IOP Conference Series: Materials Science and Engineering, vol. 53(012064), 2013.

Automated Credit Scoring System for Financial Services in Developing Countries

Rebeka Sultana, Samira Muntaha
Advanced Intelligent Multidisciplinary Systems Lab,
Department of Computer Science and Engineering,
United International University
Dhaka, Bangladesh
rebeka66_s@yahoo.com, samiramuntaha@gmail.com

Farhana Sarker
Department of Computer Science and Engineering,
University of Liberal Arts Bangladesh
farhana.sarker@ulab.edu.bd

D. M. Anisuzzaman
Department of Computer Science and Engineering,
Ahsanullah University of Science & Technology
Dhaka, Bangladesh
rajon99@gmail.com

Khondaker A. Mamun
Advanced Intelligent Multidisciplinary Systems Lab,
Department of Computer Science and Engineering,
United International University
Dhaka, Bangladesh
mamun@cse.uiu.ac.bd

Abstract—**The principle objective of this research is to carry out an intelligible evaluation of the automated credit scoring system for financial service applications. The investigation tries to identify the key determinants for constructing an automated model for credit scoring applications. It provides a comperative evaluation related to Statistical and Artificial Intelligence (AI) techniques that are used in favor of automated credit scoring. It also assists to highlight the most common and effective methods which are used in credit scoring system. Ours analysis revealed that improvements are necessary (in the existing credit scoring system) to effectively address all financial environments. Although credit scoring is highly in practice in developed countries, however in developing countries it is not implemented in many financial services. Eventhough, it has high prospect of selecting creditworthiness, but credit scoring industry is due for a major overhaul. Due to the revolution of information technology(IT) and big data it appears that new IT and big data, analytical tools can make credit scoring highly individualized, accurate, and able to provide informed decision for financial service applications.**

Keywords: Credit scoring, applications, key determinants, methods, developing countries.

I. INTRODUCTION

The incidence of lending and borrowing has related to human behavior from the Stone Age. So, credit has become substantial in daily life and it is an issue as old as trade and commerce. In general, credit refers to credit card, loans, gages, trade financing and bond etc. Credit assessment is the most troublesome effort for bank and other financial institutions. The traditional way to grant a credit card was made by human experts using prior experience and other directing logics. The 5 C's of credit, that means the character, capacity, capital, collateral and conditions are the classic methods to follow in general. Heavy training costs, incorrect and inconsistent decisions for the same applications are some common traits of this method, which may result taking a wrong decision while granting credit. As a result, financial organizations may drop creditworthy consumers or endure great capital wreck if the client eventually fail to pay [1] . Sometimes these monetary distresses can conduct to bankruptcy. These blunders have led to emergence more systematic and exact methods to assess the credit threat.

At this point, automated credit scoring has become a crucial tool for bank and other financial institutions to assess creditworthiness, reduce possibility of risks, make directional determinations, and improve the effectiveness and the financial stability. The core idea behind credit-scoring comprises the classification of potential customers into good quality applicants, who has the chance to repay the loan and bad quality applicants who has the possibility of failure to pay the loan.

Based on the literature, this study compares the effectiveness of different methods used in credit scoring system such as LDA, LR, KNN and DT and also the AI techniques such as Expert System, SVM, Fuzzy Logic, NN and GP. Moreover, it tries to highlight the main points of these methods and techniques. This work also points out some research works done in the field of credit scoring in developing countries. Furthermore, it gives some recommendations along with the possible approaches that can be taken to implement effective credit scoring in developing countries.

II. AMENITIES AND CHALLENGES OF CREDIT SCORING

A. Amenities of Credit Scoring

Credit scoring is used progressively in loan evaluation because of some obvious conveniences it possesses. As credit scoring models are based on expert systems and other artificial technology, it can make a decision very quickly as it demands less information. It reduces unnecessary variable to make a decision. Different credit experts can easily and clearly analyze the same information given the same weights which is a very important benefit of credit scoring. Scoring does the loan endorsement process very promptly. It not only saves both time and cost of the bank and customers greatly but also reduces human involvement on credit assessment Mechanized credit scoring models are correct the preference when the result histories of just acknowledged application are considering but not all applications. They do this by expecting the execution of rejected application on the off chance that they had been acknowledged [3]. The operation of the credit scoring model can be observed, followed, and balanced whenever. By the aid of credit scores, monetary organizations are able to enumerate the risks allied with yielding credit to a particular applicant quickly. The weights in the model give a measure of the relative quality of every component's relationship with credit execution. Lenders use credit scores to find out who allows for a loan along with the interest rate [4]. Automated credit scoring has a lot of amenities that accumulate to the granters as well as the suppliers. To impart a pointed scrutiny of a person's creditworthiness using scoring models, credit scores aid to minimize discrimination. This empowers credit suppliers to concentrate on just data that identifies with credit hazard and evade the individual subjectivity of a credit examiner. Enhanced objectivity in the advance endorsement procedure is another advantage of credit scoring. This objectivity aids lenders make sure for employing the equivalent underwriting benchmark to all borrowers paying little heed to race, sex, or different variables differ by law from being utilized as part of credit choices. By using credit scores, financial organizations can fix up their lending rate which they should placing their customers. Maximum-risk customers are imposed a higher lending rate. These assist financial organizations to conduct their accounts more effectively and fruitfully.

B. Challenges and Restrictions of Credit Scoring:

In spite of credit scoring has profound advantages, its few defects ought to likewise be noted. As credit scoring is a mechanical framework for investigating the advance candidate, so that there is an opportunity to break down and decipher some information inaccurately. Credit risk can never be weighed precisely, and any model that anticipates it, is erroneous. It additionally might change overnight. Example: The possessor of an industry succumbs and there is nobody qualified to supplant him. While developing a credit scoring model utilizing a one-sided example of buyers and clients who have been result credit, one of the significant issues can appear [46]. This might happen on the grounds that just the good customers are represented as the sample is one-sided and rejected customers will not be incorporated into the information for developing the model. The credit scoring system that use this example may not execute efficiently on the overall inhabitants while the record used to assemble the model is apart from the record that the model will be applied to. Therefore, if a credit scoring model has not every possible variable, usually it will a credit scoring model has every possible variable in it and it is frequently updated, normally it will be unable to classify some customers or unable to provide sufficient outcome. In statistical credit scoring, it requires a lot of data on each loan and also requires a consultant to manage and to monitor everything. It can reject faulty applications, but it cannot modify them. It is also susceptible to misuse[47]. Forces a dichotomous result, for example, either the borrower inability to pay or not is another feedback of credit scoring. Along these lines, a scope of the possible outcomes can be incorporated, as every now and again the borrower declares an issue with payments, and the loan terms can be renegotiated. In addition, credit scoring models are too excessive to purchase and prepare credit analyst furthermore fluctuate starting with one market then onto the next. Now and again a credit scoring framework might dismiss the trustworthy purchaser as a result of exchanging his/her employment or address. In spite of the confinements mentioned above, there is no hesitation that credit scoring will keep on being a noteworthy instrument in the foreseeing credit risk in consumer lending.

III. APPLICATIONS OF CREDIT SCORING

Be that as it may, utilizations of credit scoring have been generally utilized as part of various regions, together with osmosis between various factual methods utilized as a part of expectation purposes and order issues. These may be arranged into bookkeeping and finance, promoting, building and assembling, health and prescription, and general applications. However, not every one of these applications is generally utilized equally. In the early years, money related organizations utilized credit scoring basically to settle using a credit card choice for advance applications. Nonetheless, the utilization of credit scoring has developed from settling using credit card choices to settling on choices identified with lodging, protection, fundamental utility administrations, and even employment. In the field of bookkeeping and finance, financial institutions utilized scoring primarily to settle on layaway choices for loan applications. Here, credit scoring is also used for different purposes such as bankruptcy prediction and bankruptcy classification, financial distress[2], scoring applications[3] and so on. Credit scoring applications in saving money segments have extended amid the last couple of decades [4]. The assessment of new customer loan is a standout amongst the mainly essential uses of credit scoring models and has pulled in consideration over the last several decades [3]. Crediting small & medium enterprises (SME) and microfinance have been decided by credit scoring also [5]. In option to choices on individual credit applications, monetary organizations now make utilization of credit score assessments to put credit limits, oversee accessible records, and gauge the benefit of customers. Credit scoring models have additionally been utilized as a part of the protection business to settle on the uses of new protection approaches and the re-establishments of existing policies. There is considerable utilization of credit scoring in the home loan industry too [6].

IV. BASIC FACTORS OF CREDIT SCORING

The purpose of the variable selection in the credit scoring model is to obtain a role model with low dimensionality. The exactness of the model might enhance by utilizing a formal technique for picking the most suitable customer variables and the many-sided quality of the model might decrease by disposing of the non-significant variables. So, variables selection may affect the performance of the model. Predetermined scores, looked into the customer's financial record and reliability was the base to minimize the likelihood of wrongdoing and default for credit experts. A new applicant is determined by some attributes such as sex, age, marital condition, dependents, telephone, credit card, learning level, job and duration at present address. These characteristics are broadly practiced in constructing scoring models [7][8]. The working of scoring models likewise utilizes length of staying at present employment, bank account, total credit, credit duration, purpose of loan, house proprietor, car owner, month to month pay, mortgage, guarantees etc [7]. In some cases spouse's individual information, like salary, no of child, age and others has been integrated in the list of variables More variables, for example, worst record status, time in vocations, time with bank and others are less every now and again utilized as a part of building scoring models [9]. To decide individual credit scoring debt, length, credit history, payment history, types of credit and new credit are utilized. To fabricate scoring models there is no ideal number of variables that ought to be utilized. The choice of the variables contrasts from study to think about relying upon the way of information. For example,[8] applied forty-one variables, and twenty-nine variables have been utilized by[10]. Therefore, the danger system and the credit society of the organizations ought to be transformed by a part scoring model.

V. THE METHODS OF CREDIT SCORING MODEL

Credit scoring optimization is a rising topic now a day where different researchers are using different techniques for choosing the right applicant and reducing credit loss. In order to obtain a satisfactory credit scoring model, numerous methods have been proposed. In this paper, four statistical techniques have been discussed; these are DA, LR, KNN and DT. In the other hand this study also discussed five Artificial Intelligence (AI) techniques: Expert System, SVM, Fuzzy Logic, NN and GP. A short description of these techniques is discussed in this section.

A. Statistical and Optimization approaches:

Discriminant Analysis (DA) is generally used for modeling sorting tasks as a statistical technique. Fisher proposed that - DA is a classification and discrimination tool which was one of the first methods that applied to make credit scoring models by comparing between those loans which was defaulted and those which was not. DA's base assumption is that, the explanatory variables are distributed as a multivariate normal distribution with a common variance covariance matrix for each given class of response variable [25][26].

Logistic Regression (LR) is derived from linear regression. It is more suitable for fraud detection problems. Where other statistical tools failed to fit in, it can fit several kinds of distribution functions such as Gamble, Poisson, and normal distributions [27]. Another quality of this method is that, it does not need normal distribution variables and also the linearity of relationship between dependent and independent variables is not assumed in this method. The disability of LR is, it cannot properly resolve the problems of non-linear and interactive effects of explanatory variables [28].

K-Nearest Neighbor (KNN) has some fascinating features in credit scoring. For example, it is feasible to exceed the problem of population drift by using KNN, because it strongly updates by dropping older cases and by adding new candidates to the design [29][30]. But these methods have not been practiced largely in the credit scoring industry, because its predictive accuracy is extremely affected by the measure of distance and the cardinality of the neighborhood [31].

Decision Tree (DT) is a classification technique used in stimulant automated credit scoring models [32]. In order to solve the classification issues, a tree-like chart of choices and their conceivable results is used mostly. The root node of this tree is the highest node which a decision should tackle it. On an attribute or input variable, a test is done in each inward node. The leaf nodes speak to the classes and every branch taking after the node prompts the aftereffect of the test. Over fitting can be a problem of using this method.

B. Artificial Intelligence techniques:

Expert System (ES) is one of the traditional methods in accessing credit scoring. They were designed to replicate the way of thinking of human experts. In an ES the credit decision is bestow upon the local or branch lending officers [33]. In the decision making process the expert's expertise, subjective judgment and weighting of certain key factors play an important role. The advantages of using ES for credit analysis are speed and accuracy, both which far surpass human capacity.

Support Vector Machine (SVM) is a learning system that uses a linear model to map into a higher dimension feature space from the input vector using a kernel so that there is a linear reparability between the two groups [35]. Examples from the training that are close to the maximum margin hyperplane are named support vector. Normal distribution and continuity – this kind of data structures are not required which is the main advantage of SMV [39]. One of the main disadvantages of SVM is that it is sensitive to outliers or noises in the training sample due to overfitting.

Fuzzy Logic is an extension of multivalued logic. Many parameters are used for determining the credit scoring which are usually vague, difficult to define, and even conflicting.

Fuzzy set theory was developed to handle this kind of situation, and improving the accuracy of credit scoring [36]. Fuzzy rule based system provides explanation when deriving the credit score, while most of credits scoring models do not explain how the results obtained.

Neural Network (NN) contains a large number of nodes by links. By finding the complex pattern between input and output variables, NN can predict the outcome of new independent data of input. The feed-forward NN containing back-propagation (BP) is largely used for credit scoring, where the pre-layer gives signals to the neurons and output them to the next layers without feedback. The strong learning ability and no assumptions about the relationship between input variables are the main advantages of NN. Also they act as black boxes as it is difficult for humans to interpret the way neural networks reach their decision [8][42]. A disadvantage of NN is that a number of parameters like the network topology must be defined analytically.

Genetic Programming (GP) is a search heuristic that imitates the process of natural evolution [7]. Genetic Algorithms (GA) provide the solution in the form of a string. Every string is the encoded binary, real etc., version of a candidate solution. To compute a whole generation of new strings, standard GA applies genetic operators such as selection, crossover and mutation on an originally random population [40].

Table I provides a detail of different statistical approaches along with AI technologies used in various articles by the researchers. This table incorporates the analysis, references and important features of all those given methods.

TABLE I. CORERELATIVE STUDY OF DIFFERENT METHODS OF AUTOMATED CREDIT SCORING FROM PUBLISHED RESEARCH

APPROACHES	METHODS	COMMENTS
Statistical approaches (SA)	**LDA** [24][25][26]	It is still one of the most broadly established techniques to classify customers' credit score as good or bad by using linear functions. However, DA cannot properly deal with non-linear problems.
	LR [3][27][28]	It performs well on big dataset. However, this method can be applied on a small dataset or a data set with a short repayment history, but the quality of the scoring model can decrease.
	KNN [25][29][30][31]	It enables modeling of irregularities in the risk function over the feature space and a fairly intuitive procedure and can be used dynamically but its predictive accuracy is extremely affected by the measure of distance and the cardinality of the neighborhood.
	DT [32]	It solves both classification and regression problems. As like LR, it needs big dataset in order to get dependable predictions.
AI technologies	**Expert System** [33][34]	It does not end up with a score card which gives weights to each answer instead it classifies the consumers into groups, each group being homogeneous in its default risk.
	SVM [35][39][41]	It produces global optimal solution and can work well with few samples but selecting kernel and its parameters is a tricky issue.
	Fuzzy Logic [36]	It can derive human understandable rule and has low computational requirement but random choice of membership function can bias the result.
	NN [8][37][42]	It is good at function approximation, forecasting, classification, clustering and optimization tasks but demands a lot of training data and training cycles.
	GP [7][40]	It can perform better than traditional techniques such as MLP, CART, C4.5 and Rough set. However it is difficult to come out with a generic model for all class of problems. GP also requires good processing power.

Based on the literature, table II compare the effectiveness of different methods used in credit scoring system. It compares the accuracy (percentage of correctly classified instances) of the methods. The vast majority of the studies that concentrated on correlation between various strategies for credit scoring have found that artificial intelligence techniques, for example, neural networks, genetic programming and fuzzy algorithms are superior to the conventional ones taking into account the average correct classification rate criterion. However, the more straightforward classification methods, for example, LDA and LR, additionally have a decent performance in this context.

TABLE II. A COMPARISON OF DIFFERENT METHODS (SA AND AI) USED IN CREDIT SCORING SYSTEM (BASED ON THE LITERATURE).

Research Work / Methods	*West (2000)* [37]	*Lee et al. (2002)* [38]	*Baesens (2003)* [39]	*Ong et al. (2005)* [40]	*Yu et al. (2008)* [41]	*Tsai (2009)* [42]	*Chuang (2009)* [43]	*Wang (2012)* [44]
LR	81.8	73.5	79.3		73.2	84.7(avg)	76.5	71.6
LDA	79.3	71.4	79.3	80.8		76.8	76.0	
DT	77.0		77.0	78.4				69.0
NN	**82.6**	73.7 (**77.0**) hybrid LDA and NN	79.4	81.7	77.2	**92.7**	**79.5**	71.5
CART	76.9						77.5	
KNN	76.7		78.2					
GP				**82.8**				
SVM			**79.7**		**78.8**			**72.4** (avg)

VI. CREDIT SCORING IN DEVELOPING COUNTRIES

The goal of credit scoring is to measure the financial risk of the loan, so that the loan provider can make credit lending decisions quickly and objectively. Human judgment of creditworthiness can be time consuming whereas credit scoring gives advancers to find out credit worthiness in lesser time. Because of this advantage banks in developed nations as US, UK and Europe have been using credit scoring techniques with higher success rate [45]. In the developed world, they have large credit scoring firms like Equifax, Experian and TransUnion to reduce the cost of identifying creditworthy applicants. On the other hand, the lack of proper data and reliable information about the credit or monetary history of bank clients in the developing countries credit scoring can be difficult to deal with. A credit scoring system that fulfills the developing countries' need is yet to be discovered. Currently they are trying to find out an automated credit scoring technique which works best for them. In some of these countries, already have started to use credit scoring system which has been designed by developed nations and many of them are working to create their own credit scoring systems to give loan in the industrial area[15]. However, credit scoring has not been practiced effectively in small financial areas like mortgages, credit cards or personal loans. It is expected that integration of automated credit scoring system in developing countries could bring benefit to the financial sector as well as economy. As the financial organizations can determine whether there is a risk or not to grant the loan to the customer. Moreover, in developing countries small and medium enterprises (SMEs) are thought to be an important source of innovation and employment because of their flexibility in responding to new market opportunities and their potential for growth. Many developing countries have already commenced to practice automated credit scoring as a tool for their economic development.

Table III, gives an idea of adaptation of credit scoring methods and innovation being used for developing countries.

VII. PROSPECT OF CREDIT SCORING IN DEVELOPING COUNTRIES

In developing countries, financial system is mostly microfinance. Therefore, prospect in credit scoring in developing country is related to adaptation of credit scoring in microfinance.

Some recommendations for micro-lenders and microfinance institutions in developing countries:

I. The quantity of distributed credit scoring contemplates for microfinance is constrained. There is a need to broaden the geographical range of credit scoring examines towards Eastern Europe-Central Asia and Middle East-North Africa as little quantities of studies have been distributed in these regions.

II. The oppressive force execution of credit scoring frameworks for microfinance remains excessively feeble, making it impossible to legitimize a complete inversion of the conventional credit process towards scoring. However, credit scoring ought to end up a refinement instrument in the present procedure as it has effectively ended up being steady, simple to utilize, furthermore to have a specific discriminatory power. Enhancements of the discriminatory power by means of model mixes reject induction examination and more pragmatic confirmation, may bit by bit build the part of credit scoring in the credit process.

As there is no compatible credit scoring solution in developing countries, efficient approaches to implement better credit scoring system should be taken into consideration. Possible approaches that can be taken to implement effective credit scoring in developing countries are:

- Pointers of conduct got from cellphone transaction records can be prescient of loan payment [48]. To gage the prescient nature of the strategy, research has been made where joined bank information from loans have been completed with borrowers' cell telephone records. It predicts who among these people wound up repaying their loan, in light of how they utilized their cellphones before taking a loan. The examination additionally found that the prescient accuracy of the technique approaches that of credit scoring strategies utilizing conventional information as a part of more created settings.

- Involving social media in credit scoring can be a fruitful way in developing country. A huge number of people use Facebook and tracking their social media activity can help lenders calculate the risk factors. However there are negative sides of this as well. The utilization of "big data" in advertising to target particular shopper gatherings is as of now a questionable practice, for the most part since couple of customers ever acknowledges they are being tracked.

- Microfinance industry faces a big challenge on building long-term relationship with their clients. Usually, first credits are small and short-term loans. As a result, to enter the market, a special arrangement is needed. If all the terms are fulfilled by the client, they can provide clients higher amounts. It gives the client an incentive to stay with the institution.

TABLE III. SUMARRY OF RESERACH WORK THAT CONSIDER DEVELOPING COUNTRIES FOR PRACTICING AUTOMATED CREDIT SCORING

Country Name	**Adaptation of credit scoring**
Bangladesh [11]	Investigates the effect of the MFI program mediation on the moneylender interest costs in northern Bangladesh and found that moneylender financing costs increment with the rate of households borrowing getting from MFIs in the town.
India [12] [13]	They attempted to decide how far back these forecast models can anticipate that the organizations would get into financial related distress.
	Many different methods including the hybrid model of GA, Fuzzy c-means algorithm and MARS are conceptualized for prediction of bankruptcy. From the study it was visible that hybrid models work better than other static bankruptcy models.
Pakistan [14] [15]	Karachi stock exchange's non financial listed companies data were studied. Moreover it was evident that Abbas model and Altman's Z Score model was an effective tool to verify the financial stability of the company.
	To predict bankruptcy in Pakistan the most considerable financial ratios were acknowledged.
Malaysia [16]	Examine the determinants of credit hazard and demonstrated that the liquidity proportion was huge in deciding credit hazard previously, then after the fact income administration was balanced.
Iran [17] [18]	The data of diferent firmsof an organization was reviewed and a data mining model was invented to specify the non bankrupt and bankrupt firms.
	Tried to predict the failure or survival of Iranian marketplaces depending on financial ratios. In favor of this purpose GP and MDA were used.
Ghana [19]	Loan default rate in micro finance institues are still high.A fuzzy logic based approach is provided to credit scoring in order to reduce the loan default.
Nepal [20]	Demonstrated that credit risk management is an imperative indicator of bank financial execution. In this way achievement of bank execution relies on upon danger administration.
Sudan [21]	LR and DA works superior on predicting future loss.A new method has been proposed using these models to foresee bank's failure.
Vietnam [22]	Utilized a way to deal with the present shortcomings in credit scoring forms and gives a hypothetical establishment of credit scoring value.
Turkey [23]	Redesigned the quantitative analysis utilized as a part of the financial execution modules of best in class credit scoring techniques.

VIII. CONCLUSION

Credit scoring is a broadly utilized method that helps banks and other financial establishments to choose whether to allow credit to buyers who applied for loan. Nonetheless, with progression of technology, the technique for credit score should be redesigned. Utilizing enormous information to decide lower, customized rates on credit cards and loans will advantage monetarily dependable individuals in a way the present framework does not, permitting reliable borrowers to pay less and escape obligation speedier. Also, fabricating a client base of monetarily mindful individuals will advantage banks - decreasing the dangers of misrepresentation, and additionally default and sparing organizations cash over the long haul. Putting resources into the fate of stable people similarly puts resources into the fate of a stable financial industry and national economy. It is the ideal opportunity for the financial framework to grasp a bigger extent of markers to decide budgetary obligation; basically, the old methods for assessment can not stay aware of better approaches for living and working. This begins with social affair a superior comprehension of buyers to make financial arrangements that fit their individual needs. Instruments are being created that will make lending more productive to both the banks and the borrowers. Equipped with capable software and refined data science, the eventual fate of the financial industry lays on the capacity to bring educated individual finance into the current period.

References

[1] M. C. Chen and S. H. Huang, "Credit scoring and rejected instances reassigning through evolutionary computation techniques," *Expert Syst. Appl.*, vol. 24, no. 4, pp. 433–441, 2003.

[2] Y.-C. Hu, "Incorporating a non-additive decision making method into multi-layer neural networks and its application to financial distress analysis," *Knowledge-Based Syst.*, vol. 21, pp. 383–390, 2008.

[3] J. N. Crook, D. B. Edelman, and L. C. Thomas, "Recent developments in consumer credit risk assessment," *Eur. J. Oper. Res.*, vol. 183, no. 3, pp. 1447–1465, 2007.

[4] D. Hand, S. Sohn, and Y. Kim, "Optimal bipartite scorecards," *Expert Syst. Appl.*, vol. 29, no. 3, pp. 684–690, 2005.

[5] J. Gupta, N. Wilson, A. Gregoriou, and J. Healy, "The effect of internationalisation on modelling credit risk for SMEs: Evidence from UK market," *J. Int. Financ. Mark. Institutions Money*, vol. 31, no. October, pp. 397–413, 2014.

[6] A. Haughwout, R. Peach, and J. Tracy, "Juvenile delinquent mortgages: Bad credit or bad economy?," *J. Urban Econ.*, vol. 64, no. 2, pp. 246–257, 2008.

[7] C. Ong, J. Huang, and G. Tzeng, "Building credit scoring models using genetic programming," *Expert Syst. Appl.*, vol. 29, no. 1, pp. 41–47, 2005.

[8] M. Leshno and Y. Spector, "Neural network prediction analysis: The bankruptcy case," *Neurocomputing*, vol. 10, no. 2, pp. 125–147, 1996.

[9] T. Bellotti and J. Crook, "Support vector machines for credit scoring and discovery of significant features," *Expert Syst. Appl.*, vol. 36, no. 2, pp. 3302–3308, 2009.

[10] L. M. Salchenberger, E. M. Cinar, and N. a. Lash, "Neural Networks: A New Tool for Predicting Thrift Failures," *Decis. Sci.*, vol. 23, no. 4, pp. 899–916, 1992.

[11] D. Mallick, "Microfinance and Moneylender Interest Rate: Evidence from Bangladesh," *World Dev.*, vol. 40, no. 6, pp. 1181–1189, 2012.

[12] M. a, V. Gayathri, and E. Al, "a Hybrid Model for Bankruptcy Prediction Using Genetic," *Int. J. Soft Comp*, vol. 2uting (IJ, no. 1, pp. 12–24, 2011.

[13] R. Bott, "No Title No Title," *Igarss 2014*, vol. 7, no. 1, pp. 1–5, 2014.

[14] A. Rashid and Q. Abbas, "Predicting Bankruptcy in Pakistan," *Theor. Appl. Econ.*, vol. 18, no. 9, pp. 103–128, 2011.

[15] M. S. Roomi, W. Ahmad, M. Ramzan, and M. Zia-ur-Rehman, "Bankruptcy Prediction for Non-Financial Firms of Pakistan," *Int. J. Account. Financ. Report.*, vol. 5, no. 2, p. 26, 2015.

[16] N. A. Manab, N. Y. Theng, and R. Md-Rus, "The Determinants of Credit Risk in Malaysia," *Procedia - Soc. Behav. Sci.*, vol. 172, pp. 301–308, 2015.

[17] M. M. Shiri and M. Ahangary, "Corporate Bankruptcy Prediction using Data Mining Techniques: Evidence from Iran," *African J. Sci. Res. Vol*, vol. 8, no. 1, 2012.

[18] H. Etemadi, A. A. Anvary Rostamy, and H. F. Dehkordi, "A genetic programming model for bankruptcy prediction: Empirical evidence from Iran," *Expert Syst. Appl.*, vol. 36, no. 2 PART 2, pp. 3199–3207, 2009.

[19] U. F. Ibn, J. K. Panford, J. Ben Hayfron-, and R. Base, "Fuzzy Logic Approach to Credit Scoring for Micro Finances in Ghana (A Case Study of KWIQPLUS Money Lending)," vol. 94, no. 8, 2014.

[20] R. Prakash and S. Poudel, "The impact of credit risk management on financial performance of commercial banks in Nepal," *Int. J. Arts Commer.*, vol. 1, no. 5, pp. 9–15, 2012.

[21] Khalafalla Ahmed Mohamed Arabi, "Predicting Banks' Failure: The Case of Banking Sector in Sudan for the Period (2002-2009)," *J. Bus. Stud. Q.*, vol. 4, no. 3, p. 160, 2013.

[22] V. O. Hong and N. Dinh, "A new approach to determining credit rating & its applications to vietnam 's listed firms," *42nd Aust. Conf. Econ.*, no. April 2013, pp. 1–26, 2013.

[23] A. B. Emel, M. Oral, A. Reisman, and R. Yolalan, "A credit scoring approach for the commercial banking sector," *Socioecon. Plann. Sci.*, vol. 37, pp. 103–123, 2003.

[24] E. I. Altman, "Financial ratios, discriminant analysis and the prediction of corporate bankruptcy," *J. Finance*, vol. XXIII, no. September, 1968.

[25] I.-C. Yeh and C. Lien, "The comparisons of data mining techniques for the predictive accuracy of probability of default of credit card clients," *Expert Syst. Appl.*, vol. 36, no. 2, pp. 2473–2480, 2009.

[26] R. a. Eisenbeis, "Problems in applying discriminant analysis in credit scoring models," *J. Bank. Financ.*, vol. 2, pp. 205–219, 1978.

[27] H. Zhao, A. P. Sinha, and W. Ge, "Effects of feature construction on classification performance: An empirical study in bank failure prediction," *Expert Syst. Appl.*, vol. 36, no. 2, pp. 2633–2644, 2009.

[28] B. Baesens, T. Van Gestel, S. Viaene, M. Stepanova, J. Suykens, and J. Vanthienen, "Benchmarking state-of-the-art classification algorithms for credit scoring," *J. Oper. Res. Soc.*, vol. 54, no. 6, pp. 627–635, 2003.

[29] R. S. Society, "for assessing consumer credit risk classifier A k-nearest-neighbour," vol. 45, no. 1, pp. 77–95, 2014

[30] I. Brown and C. Mues, "An experimental comparison of classification algorithms for imbalanced credit scoring data sets," *Expert Syst. Appl.*, vol. 39, no. 3, pp. 3446–3453, 2012

[31] M. Keynes and M. Keynes, "Construction of a k-nearest-neigbbour credit-scoring systemf," pp. 305–321, 1997.

[32] T.-S. Lee, C.-C. Chiu, Y.-C. Chou, and C.-J. Lu, "Mining the customer credit using classification and regression tree and multivariate adaptive regression splines," *Comput. Stat. Data Anal.*, vol. 50, no. 4, pp. 1113–1130, 2006.

[33] A. Bahrammirzaee, "A comparative survey of artificial intelligence applications in finance: artificial neural networks, expert system and hybrid intelligent systems," *Neural Comput. Appl.*, vol. 19, no. 8, pp. 1165–1195, 2010.

[34] K. Bryant, "ALEES: an agricultural loan evaluation expert system," *Expert Syst. Appl.*, vol. 21, no. 2, pp. 75–85, 2001.

[35] T. Bellotti and J. Crook, "Support vector machines for credit scoring and discovery of significant features," *Expert Syst. Appl.*, vol. 36, no. 2 PART 2, pp. 3302–3308, 2009.

[36] Y.-R. Syau, H.-T. Hsieh, and E. S. Lee, "Fuzzy Numbers in the Credit Rating of Enterprise Financial Condition," *Rev. Quant. Financ. Account.*, vol. 17, no. 4, pp. 351–360, 2001.

[37] D. West, "Neural network credit scoring models," vol. 27, 2000.

[38] T.-S. Lee, C.-C. Chiu, C.-J. Lu, and I.-F. Chen, "Credit scoring using the hybrid neural discriminant technique," *Expert Syst. Appl.*, vol. 23, no. 3, pp. 245–254, 2002.

[39] J. Crook, "Developing intelligent systems for credit scoring using machine learning techniques," no. 180, 2003.

[40] C. Ong, J. Huang, and G. Tzeng, "Building credit scoring models using genetic programming," *Expert Syst. Appl.*, vol. 29, no. 1, pp. 41–47, 2005.

[41] L. Yu, S. Wang, and K. K. Lai, "Credit risk assessment with a multistage neural network ensemble learning approach," vol. 34, pp. 1434–1444, 2008.

[42] M. Tsai, S. Lin, C. Cheng, and Y. Lin, "Expert Systems with Applications The consumer loan default predicting model – An application of DEA – DA and neural network," *Expert Syst. Appl.*, vol. 36, no. 9, pp. 11682–11690, 2009.

[43] C. Chuang and R. Lin, "Constructing a reassigning credit scoring model," *Expert Syst. Appl.*, vol. 36, no. 2, pp. 1685–1694, 2009.

[44] G. Wang and J. Ma, "Expert Systems with Applications A hybrid ensemble approach for enterprise credit risk assessment based on Support Vector Machine," *Expert Syst. Appl.*, vol. 39, no. 5, pp. 5325–5331, 2012.

[45] V. Tech, "Credit-scoring models in the credit-onion environment using neural networks and genetic algorithms," pp. 323–346, 1997.

[46] H. Koh, W. Tan, and C. Goh, "A two-step method to construct credit scoring models with data mining techniques," *Int. J. Bus. …*, vol. 1, no. 1, pp. 96–118, 2006.

[47] M. Schreiner, "Benefits and Pitfalls of Statistical Credit Scoring for Microfinance Author ' s Note Benefits and Pitfalls of Statistical Credit Scoring for Microfinance 1 . What is scoring ?," vol. 28, no. 1, pp. 63–86, 2004.

[48] D. Björkegren and D. Grissen, "Behavior Revealed in Mobile Phone Usage Predicts Loan Repayment," pp. 1–10, 2015

An Investigation of the Various Methods of Lip Reading Systems

Muhammad Mamunur Rashid
Department of Computer Science and Engineering
Port City International University, Chittagong, Bangladesh
E-mail: mmr.engr@gmail.com

Rashed Mustafa
Department of Computer Science and Engineering
University of Chittagong, Bangladesh
E-mail: rashed.m@cu.ac.bd

Mohammad Sanaullah Chowdhury
Department of Computer Science and Engineering
University of Chittagong, Bangladesh
E-mail: sana1691@gmail.com

***Abstract*— There are continuous research and debate on lip reading specially how to get accurate speech recognition. This paper presents and analyzes various lip reading techniques to classify the discrete utterance with acoustic and without acoustic signal based on visual feature analysis to improve the highest efficiency of speech recognition. This Paper also describes feature extraction methods, applications of different environment, highest accuracy, limitations, numeric experimental comparisons and future research direction.**

Keywords –Lip-reading, visual speech reading, visual feature extraction, SVM, KNN, optical flow, HMMs etc.

I. INTRODUCTION

Recognizing speech in noise or noiseless environment is one of the major research challenges in real world and machine vision [1]. Speech perception is multimodal process which is not involved only audio information but also visual information. Although we can't use only speech hear (audio) or only visual information for getting highest performance of speech recognition [2]. Both combination can be improved the performance of speech recognition as well as used in different methods. If there are limited numbers of utterance with lip shape to be identified then it can be possible to use as visual information to do speech recognition. This phenomenon is widely known as lip reading [3]. It is an important alternate to traditional speech recognition technology. It can be used whether audio is seriously corrupted or unavailable by background noise.

For a practical point of view lip reading is considered as subject depended (SD) [24, 30] and Subject independent (SI) systems [25, 30]. The SD system can be used in private environment and customized for a particular user. For example in a smart vehicle, read voice command from the driver to operate different devices. Earlier the training data is provided directly by the user.

On the other hand, the SI system is used to serve large crowd users in an open environment. For instance a social robot can be captured vocal querries from a user in a public environment. To improving the highest speech recognition [12] via lip reading, there are several techniques/methods used, such as:

A. Active appearance models. (AMMs)[26].
B. Using Classifier based phoneme [27].
C. Feature extraction from frequently and infrequently nature of changing images [28].
D. Combination of geometric and appearance based features [29].
E. Image space for temporal interpolation [30].
F. Motion estimation of lips using optical flow [31]

The figure 1 shows basic block diagram of lip reading system.

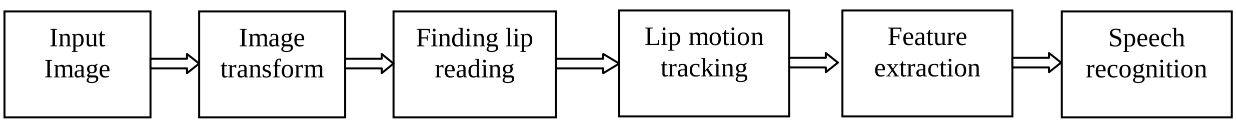


Figure 1: Block Diagram of Lip Reading System [28]

The rest of this paper is organized according to the following structure: section II describes a brief literature on traditional lip reading systems, a comparative analysis is demonstrated in section III, section IV presents a comprehensive discussion and section V concludes the paper.

II. LIP READING SYSTEMS

A. Active Appearance Models (AAMs)

An Active appearance model [26] is a model that consists of a region of lip shape plus an appearance component of frame from a video [1]. The shape frame is constructed a set of images with x and y co-ordinates in a mesh with a set of vertices and applying major element of lip region. This technique uses the following two tools.

Hierarchical LinearDiscriminant Analysis (HiLDA)

HiLDA (Hierarchical Linear Discriminant Analysis) is used to capture the dynamic of visual speech and also used to reduce the dimensionally of feature vectors [2]. To compute the features of HiLDA, linear discriminant (LDA) is applied in high dimensional feature vector formed by concatening 1^{st}, 2^{nd} and 3^{rd} frames in a 15-frame window. After that a MLLT [3] is applied within class distance to minimize and maximize of feature map.

Kaldi Pipeline and DNN for Speech Rrecognition:

Kaldi is a toolkit for speech recognition with DNN (Deep neural Network) and ANN with several hidden layers of the input and output units. All first raw features are required to be prepared and converted into a suitable format for reading and writing in Kaldi . The 39 dimensional mel frequency cepstral coefficient (MFCCs) [26] with energy of the acoustic feature delta-delta and delta co-efficient are appended. The 23-dimensional vectors are formed by concatenating the audio and visual features. Decoding system consists of three steps such as- courage model rescoring decode, lattice generation and decoding [20] graph generation. In this method DNN gave 85% word accuracy which is huge improvement on the baseline HMM performance.

WFST Decoder:
The three steps are used to decode the speech utterances; to decoding graph generation, lattice and decode generation and tongue model rescoring. The graph of decoding using either a GMM-HMM or a DNN-HMM is modeled as a finite-state transducer (FST) formed by composing an FST symbol of the HMM, pronunciation dictionary, phonetic context-dependency and tri-gram language model.

$$HCLG = \min(\det(HoC\ oLoG))$$

Where ,

H= WFST composition of HMM structure
C= phonetic context-dependency
L=Lexicon
G= language model or grammar .

B. Using Classification based Phoneme

In this method visems [27] is dependent on speaker. This is a visual cue which is representative of a subset of phonemes on lips [4, 5, and 6]. So a set of visemes classifier are inherently smaller that a set of phoneme classifiers. The unit of acoustic speech is called phones and the unit of visual speech is called visems. The relation between phonemes to viseme is called P2V maps [4]. A set of confused matrices for each speaker used to cluster together. It is performed on word recognition using classifier based phoneme labels. In the clustering process a new P2V mapping is reclassified of a new viseme grouping each time of phoneme [27]. This process is 45 times P2V maps per speaker. New classifiers are then used to repeat word recognition task.

There are three steps are used for viseme recognition

- Phoneme recognition
- SD phoneme clustering
- Viseme recognition.

Phoneme recognition:

For phoneme, single-state tied short pause or HMM short silences are used between words in the uttering sentences. There are also used bigram word network to support recognition.

SD phoneme clustering:

Phonemes are assigned to vowels and consonants. Vowels and consonants cannot be mixed and pair score be merged equal a score be broken randomly. This process is repeated for a possible set of visual units.

Viseme recognition:

It is based on Gaussian mixture of time components and three state HMMs. A single state-tied short –pause is used between words of uttering sentences. There are also used a bigram word network to support recognition of word time around visime classes to be used as recognition.

Word recognition measurement:
The Word recognition performance to the HMMs is measured correctness of A and C

$$C= \frac{N-D-S}{N}$$

$$A= \frac{C-I}{N}$$

Where , S means number of substitutions error, D means number of deletion error, I means insertions errors, N means total number of labels.

C. Feature extraction from Frequently and Infrequently nature of Changing Images

In this method, little common scene detection of video techniques HD, ECR, SCDSW and SAD are elucidated. The proposed method is the feature extraction from changes of scenes on a ground truth of frequently or infrequently nature of changing images [28].

Method:

At first ffmpeg [7] (open source tool) is to extract all key frames. Then the result is summarized for key frames of external videos.A set of feature are extracted from the image. The features are classified. The statics methods analyze is use to classify by the spatial distribution of gray values by computing local feature at every point in the image. Then the texture is described for statistic measures. The statistical methods are further classified into one pixel (first order), two pixel (2^{nd} order), then three or more pixel (higher-order) statistics [8]. The available used of statistical methods are gray level co-occurrence matrix(GLCM), Histogram properties, auto correlation, gray level Run-length(GLRL) and Local Binary pattern(LBP). Due to the relatively small size of image and large number of images in data set ordinary histogram is suitable for simplify and calculation. There are eight feature are considered to classify in the method such that minimum, maximum, mode gray value, variance, mean, skewness, kurtosis and standard deviation [28].

By using these features, considering time sequence between continuous frames are used for word recognition.

The KNN:

The k-nearest neighbor algorithm (KNN) is a technique to classifying objects based on closet training examples in the feature space in pattern recognition. KNN is a method of supervised and instance-based education or idle learning wherever the purpose is only approximated nearby and all computation is deferred in anticipation of classification. The KNN algorithm is amid the simplest of all machine learning algorithms: where an object is classified through a mass vote of its neighbors, through the object being assigned to the class the majority common amongst its k nearest neighbors (k is a positive integer, typically small). If k = 1, after that the object is easy assigned to the class of its adjacent neighbor [39]. The KNN algorithm is very simple and easy to execute. There is not an explicit model it. And the KNN is mainly effective technique for the type of data variables including the characteristics. In the classification phase, k is a user-defined steady, and an unlabeled vector (a query or test spot) is classified by assigning the label which is the majority frequent among the k training samples nearest to that query spot [28].

D. Combination of geometric and appearance based features

This method describes the uses of visual feature describing geometry and appearance [29] of the lips. Also provide a reasonable tradeoff between accuracy and degeneration of the models. There are use HMM (Hidden Markov Model) to implemented of isolated digit recognition and the phonetic boundaries are used to estimate the video sequence for each digits. The CSIRO toolkit [29] implements mode fitting by regularized to align the facial expression. After that the facial land marks is detected in the frames automatically. For face detection viola-Jones face detector is used. In the part of data processing the recordings are segmented into turns to align the speech signal the open source SAILAlign [9] is used. Also calculate for each digit the phonetic boundaries are used to estimate in video sequence. This method also implemented a minimize error on extraction of facial land-marks after normalization [10].

The HMM:

Hidden Markov models (HMMs) are a formal foundation for making probabilistic models of linear sequence 'labeling' problems They provide a conceptual toolkit for building complex models just by drawing an intuitive picture. They are at the heart of a diverse range of programs, including genefinding, profile searches, multiple sequence alignment and regulatory site identification. HMMs are the Legos of computational sequence analysis.

E. Image space for temporal Interpolation

This system has been developed to separate strategies for SD (Subject depended) and SI (Subject In depend) lip reading. To perceive speech a multimodal process is required, such that information of hearing and visual [11].The combination of hearing and visual information, it could significant improvement of speech recognition [12]. Image space for temporal interpolation [30] is a simple curve matching problem to tackle the STSS problem; it has projected a video normalization method to get better the system. It is suitable for lip reading system can be considered as SD and SI system.SD is used in private environment, like inside car, such a system may perhaps be build in a smart means of transportation to operate different device by voice command.

In the system, instruction data is probably provided straight by the user. On the other hand, a subject in depend system is used to work a large group of work users, like public places. Example is that a social robot can be capture vocal queries from large number of users [30].

The SD system faces various challenge in environment like that user to access system including large variation within lip shapes , skin texture around the mouth, varying speed and different accent which has occurred significant effect the spatiotemporal appearance of a speaking mouth.

Support vector machine (SVM) classifier:

Support vector machine (SVM) classifier intended for linear and separate container is able to get the best separating hyper plane between classes inside sparse high-dimensional spaces by means of comparatively a small number of training data. SVM maximizes the expanse of the sorting out hyper plane from the neighboring training data point called the support vectors [30]. SVMs were chosen outstanding to the talent of SVMs to discover a globally best decision function to split the different classes of data. The LIBSVM toolbox (Chang & Lin, 2001) [30] is used in the experiments to plan the c-SVMs. The one-vs-all multi-class SVM method is adopted in the training of SVM classifier. One SVM was formed to study each viseme. The gamma parameter and the mistake expression penalty parameter C, of the kernel function are optimized using iterative experiments (grid search).

Local Binary Pattern **(**LBP)**:**

LBP is a straightforward yet very well-organized texture operative which is labeled the pixels of an image through thresholding the neighborhood of every pixel and considers the effect as a binary number. Outstanding to its discriminative power and computational cleanness, LBP texture operator is going to a admired advance in a variety of applications. LBP be able to be seen as a unifying move toward to the conventionally different geometric and structural models of surface study. Possibly the majority significant possessions of the LBP operator in real-world applications that toughness to monotonic gray-scale changes caused, used for example, through light variations. An additional important assets is its computational is very simplicity, which makes LBP feasible to analyze images in demanding real-time settings in the world.

Multiple kernel learning **(**MKL):

MKL **is** a set of machine learning methods that make use of a predefined set of kernels and study an best linear or non-linear grouping of kernels as element of the algorithm. The reasons to make use of multiple kernel learning contain a) capability to choose for an optimal kernel and parameters from a superior set of kernels, dropping bias due to kernel collection as allowing for additional automatic machine learning methods, and b) the combining data from dissimilar sources (e.g. echo and images from any video) that have dissimilar ideas of parallel and therefore have need of dissimilar kernels.As an option of creating a new kernel, many kernel algorithms be able to be used to merge kernels by now recognized for each entity data supply.

F. Optical flow analysis for motion of lips **.**

In this method represent a lip reading technique to analysis the discrete utterance without measuring acoustic signal [31]. The technique analysis video data of lips motions by computing optical flow and using SVM. Inter and intra speed of speech chances can beat through normalization by using interpolation and MSE. Speech based system is useful only a natural information can be used in a system there are few number of systems proposed that mechanical sensing of facial movement, palate's movement, recording facial muscle activities, facial and measuring intra oral pressure [13].Kumar et. al. [14] has reported on a speech recognition based on electromyography (EMG) signals of speaker's facial muscles. The VSR scheme can be largely categorized in two system, such as- shape-based [15] and pixel-based approaches. Shape-based approaches are used to identify contour of mouth for classifying utterances. Pixel based feature are used for image transforms like as DCT [15][16][17][18]. Discrete wavelet transform (DWT) [18] principal component analysis and linear discriminate analysis (LDA)[16].LDA is used for lip reading. The other feature extraction methods are that analysis of image sequence to represent dynamics of the lip while speaker spooked. Yan et al[19] propose for dynamic lip reading approach using MHI or spatio Temporal Templates .The One general problem in audio visual speech recognition system is sensitivity of the system to the inter-speaker variations[31]. For that subject in depend application are not supported by the systems.

Proposed technique in optical flow analysis for motion of lips:

The method consists of automatic temporal segmentation (i.e. Compute start and stop frames of an utterance) of isolated utterances (achieved by pair wise pixel distance) which called temporal matching. Then converted the color images to binary images then mean difference of accumulative frames, and then squared to get absolute values and prominent values. Then resultant signal are smoothed using linear interpolation method and further smoothed by Gaussian filtering [31].

In the final step after evaluating the start and end frame of an utterance fed into VSR system. The system extracted VSF (visual speech features from the image data that were fed into SCM system to classify. In feature extraction zero energy difference frames are filtered out using MSE method. Then it reduces computational burden while calculation optical flow (optical flow [31] is the system which is measured of visually apparent motion of object in video data. It also measure spatio-temporal variance between two image frames).

Result calculation:

The result is calculated [31] by

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \text{ x } 100\%$$

$$\text{Sensitivity} = \frac{Tp}{TP+FN} \text{ x } 100\%$$

$$\text{Specificity} = \frac{TN}{FP+TN} \text{ x}100\%$$

Where, TP= True Positive, TN=True Negative,FP= False Positive, FN= False Negative

III. COMPARISON OF VARIOUS METHODS OF LIP READING SYSTEMS

<table>
<tr><th>Sl</th><th>Method Name</th><th>Accuracy, specificity, sensitivity other specifications</th><th>Remarks</th></tr>
<tr><td>A</td><td>Active appearance model(AMMs)[26]</td><td>
<table>
<tr><th>Decoder</th><th>features</th><th>Models</th><th>SE</th><th>%Word Acc</th></tr>
<tr><td rowspan="4">WFST [Khaldi]</td><td>AAM</td><td>CD-DNN</td><td>.36</td><td>77.49</td></tr>
<tr><td>HilDA</td><td>CD-DNN</td><td>.30</td><td>84.67</td></tr>
<tr><td>AAM</td><td>CD-GMM</td><td>.81</td><td>49.19</td></tr>
<tr><td>AAM</td><td>CI-GMM</td><td>.80</td><td>37.76</td></tr>
<tr><td rowspan="2">HTK</td><td>AAM</td><td>CI-HMM</td><td>.38</td><td>33.32</td></tr>
<tr><td>AAM</td><td>CD-HMM</td><td>.34</td><td>47.48</td></tr>
</table>
</td><td>Comparing between baseline System of only visual speech recognition. Proposed system's result is shown using DNN on WFST decoder. (CI-Context independent and CD- Context dependent)</td></tr>
<tr><td>B</td><td>Using Classifier based phoneme[27]</td><td>Poor performance while vesime is less than 10 and optimum performance for 10-20 vesimes.
Word recognition is measured by correctness of used classifier.</td><td>HMMs[22] is used as Classifier</td></tr>
<tr><td>C</td><td>Feature extraction from frequently and infrequently nature of changing images[28]</td><td>The classification accuracy is depending on noise. Noise is reduced by the large values of K. Various heuristic techniques is used for a good K.
(e.g: cross validation)</td><td>K is positive integer and typically small , KNN[21] is used as Classifier.</td></tr>
<tr><td>D</td><td>Combination of geometric and appearance based features[29]</td><td>
<table>
<tr><th>Test</th><th>Train</th><th>SD (%)</th><th>SI (%)</th><th>ADPT (%)</th></tr>
<tr><td>WHI</td><td>NEU</td><td>71.85</td><td>50.87</td><td>68.14</td></tr>
<tr><td>NEU</td><td>NEU</td><td>80.78</td><td>52.93</td><td>77.31</td></tr>
<tr><td>WHI</td><td>WHI</td><td>82.64</td><td>52.34</td><td>76.24</td></tr>
</table>
Data is showed that accuracy of digit recognition on
SD-Speaker dependent
SI- Speaker Independent
ADPT-Model adaption</td><td>WHI- Whisper Speech[23]
NEU- Neutral speech

HMM system is used as classifier.</td></tr>
<tr><td>E</td><td>Image space for temporal interpolation[30]</td><td>
<table>
<tr><th>SD (Result)</th><th>noisy</th><th>clean</th></tr>
<tr><td>$LBP^{u2}_{(16,4)}$</td><td>83.3%</td><td>96.3%</td></tr>
<tr><td>$LBP^{u2}_{(8,3)}$</td><td>81.5%</td><td>95.2%</td></tr>
<tr><td>$LBP^{u2}_{(16,4)} + LBP^{u2}_{(8,3)}$</td><td>85.1%</td><td>96.5%</td></tr>
<tr><td>Zhou et al.[18]</td><td>64.2%</td><td>n/a</td></tr>
<tr><td>Zhou et al.[19]</td><td>n/a</td><td>90.7%</td></tr>
<tr><th>SI (Result)</th><td colspan="2">Recognition rate on noisy data</td></tr>
<tr><td>$LBP\text{-}TOP^{1x5x3}_{8,3}$ with normalization SVM</td><td colspan="2">76.7%</td></tr>
<tr><td>$LBP\text{-}TOP^{1x5x3}_{8,3}$ Without normalization [18]</td><td colspan="2">58.6%</td></tr>
<tr><td>MKL-Fusion with normalization SVM</td><td colspan="2">81.3%</td></tr>
</table>
</td><td>SD- Subject Dependent[24

SI-Subject independent[25]
LBP- Local Binary Pattern
MKL- Multiple kernel Learning.

Used classifier is SVM</td></tr>
<tr><td>F</td><td>Motion estimation of lips using optical flow[31]</td><td>
<table>
<tr><th></th><th>Result</th></tr>
<tr><td>Accuracy</td><td>95.9%</td></tr>
<tr><td>Specificity</td><td>98.1%</td></tr>
<tr><td>Sensitivity</td><td>66.4%</td></tr>
</table>
</td><td>System is suitable for noisy environment SVM is used as Classifier</td></tr>
</table>

IV. Discussion

Active Appearance Model (AMMs) - AAMs consist of shape component of the lip section in a video frame. The lip contour is constructed x and y co ordinates of the vertices of a mesh. In this method used HiLDA features are used to get dynamics image speech. The Maximum Likelihood Linear Transform is applied to diagram feature. Minimize within class distance and maximize between class distances. In this method DNNs proved 85% work accuracy, DNNs improved the robustness audio visual and audio-only tasks.

Using Classifier based phoneme- In This method made a P2V mapping for each time when a phoneme is re-classified. There are deriving up to 45 P2V maps per speaker. In the method used 3 steps: Phone recognition, Speaker dependent phoneme clustering, viseme recognition. The Class feature are normalize by subtracting corresponding mean from each feature frame across the video using HTK toolkit for HMMs. This method also is compared viseme sets with a controlled manner.

Feature extraction from frequently and infrequently nature of changing images – The method identify feature from frequently nature or infrequently nature of image in gray scale by measuring mean value, variance of gray scale, skewness, maximum gray scale, minimum gray level value, mode value (Kurtosis) To extract the feature, histogram feature is used as statistical feature which analyze spatial distribution of gray values.

The eight techniques are applied to extract feature for a gray scale. The key frame is used for face detection with ROI when people utter a word. Also consider time sequence between frames for recognizing word.

Combination of geometric and appearance based features- This method discussed a reasonable tradeoff between geometric and appearance based feature of visual information using HMM. To extract feature HMMs approach is used which combine geometric and appearance based features. This methods approach lip reading in Speaker depended system accuracy is 80.78%. It is the feasible alternate to improve the performance of whisper speech recognition

Image space for temporal interpolation - In this method SD and SI system are used. SD system can model and a map high dimensional feature is extracted onto low dimensional continuous features curve with determining a set of trigonometric function. LBP descriptor is used to calculate LBP feature on whole video sequence as a volume and accumulated temporal feature which normalize LBP histogram. Method proposed individual technique for SD and SI lip reading. It is turned into a plain curve identical problem to measure STSS problem for SD and in SI system used a video.

Motion estimation of lips using optical flow-This method is described that the dynamics visual speech features named as Motion History Image or STT. The MHI is a gray level image which occurred speech articulators in image sequence when and where movement. Optical flow analysis is used to extract feature from videos.

This method is based on motion capturing of lips by using optical flow and featured are classified by SVM. This technique is produced overall accuracy of 95.95%, Specificity 98.1% and Sensitivity 66.4%. It can be used drive computerized machinery in noisy environment also can be applied to impaired people to rehabilitation.

V. Conclution

This study investigated six methods to recognize speech on the basis of lip reading with extracting feature of visual. Most of methods are described how to maximize word accuracy, specificity, sensitivity other specifications with feature extraction methods and suggest future work with necessary use in different environment. A comprehensive comparison is also showed the outcome of the works in section III.

References

[1] Price, Patti, et al. "The DARPA 1000-word resource management database for continuous speech recognition." *Acoustics, Speech, and Signal Processing, 1988. ICASSP-88., 1988 International Conference on*. IEEE, 1988.

[2] Potamianos, Gerasimos, Juergen Luettin, and Chalapathy Neti. "Hierarchical discriminant features for audio-visual LVCSR." *Acoustics, Speech, and Signal Processing, 2001. Proceedings.(ICASSP'01). 2001 IEEE International Conference on*. Vol. 1. IEEE, 2001.

[3] Gopinath, Ramesh A. "Maximum likelihood modeling with Gaussian distributions for classification." *Acoustics, Speech and Signal Processing, 1998. Proceedings of the 1998 IEEE International Conference on*. Vol. 2. IEEE, 1998.

[4] Bear, Helen L., Richard W. Harvey, Barry-John Theobald, and Yuxuan Lan. "Which Phoneme-to-Viseme Maps Best Improve Visual-Only Computer Lip-Reading?." In *Advances in Visual Computing*, pp. 230-239. Springer International Publishing, 2014.

[5] T. Chen and R. R. Rao, "Audio-visual integration in multimodal communication," Proceedings of the IEEE, vol. 86, no. 5, pp. 837–852, 1998.

[6] C. G. Fisher, "Confusions among visually perceived consonants,"Journal of Speech, Language and Hearing Research, vol. 11,no. 4, p. 796, 1968.

[7] TF Cootes, GJ Edwards, CJ Taylor. *Active Appearance Models.* Proc. European Conf. Computer Vision. 1998

[8] Unay, Devrim: "Robustness of local binary patterns in brain MR image analysis." *Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE*. IEEE, 2007.

[9] Qin Jin, Tanja Schultz, and Alex Waibel. "Far-field Speaker Recognition". *In IEEE transactions on Audio, Speech, and Language Processing (TASL), Vol. 15, No. 7, September, 2007.*

[10] Katsamanis, Athanasios, et al. "SailAlign: Robust long speech-text alignment." *Proc. of Workshop on New Tools and Methods for Very-Large Scale Phonetics Research*. 2011.

[11] H. McGurk and J. MacDonald. Hearing lips and seeing voices. Nature, 264:746–748, 1976.

[12] G. Potamianos, C. Neti, and G. Gravier. Recent advances in the automatic recognition of audio-visual speech. Proceedings of the IEEE, 91(9):1306–1326, 2003.
[13] Soquet, Alain, Marco Saerens, and Véronique Lecuit. "Complementary cues for speech recognition." *Proc. Int. Conf. Phonetic. Sci*. 1999.
[14] Kumar, Sanjay, et al. "EMG based voice recognition." *Intelligent Sensors, Sensor Networks and Information Processing Conference, 2004. Proceedings of the 2004*. IEEE, 2004.
[15] Potamianos, Gerasimos, et al. "Recent advances in the automatic recognition of audiovisual speech." *Proceedings of the IEEE* 91.9 (2003): 1306-1326.
[16] Mase, Kenji, and Alex Pentland. "Automatic lipreading by optical-flow analysis." *Systems and Computers in Japan* 22.6 (1991): 67-76.
[17] Pérez, Jesús F. Guitarte, et al. "Lip Reading for Robust Speech Recognition on Embedded Devices." *ICASSP (1)*. 2005.
[18] Lucey, Patrick, and Gerasimos Potamianos. "Lipreading using profile versus frontal views." *Multimedia Signal Processing, 2006 IEEE 8th Workshop on*. IEEE, 2006..
[19] W. C. Yau, D. K. Kumar, and S. P. Arjunan. ”Visual speech recognition using dynamic features and support vector machines, International Journal of Image and Graphics, vol.8, pp. 419-437, 2008.
[20] K. Vertanen, “Baseline WSJ acoustic models for HTK and Sphinx: Training recipes and recognition experiments,” Cambridge, United Kingdom: Cavendish Laboratory, 2006
[21] K-nearest neighbor algorithm, available at, http://en.wikipedia.org/wiki/K-nearest_neighbor_algorithm
[22] Evermann, Gunnar, et al. *The HTK book*. Vol. 2. Cambridge: Entropic Cambridge Research Laboratory, 1997 (accessed at: 3/2/2016.
[23] Tran, Thomas, Soroosh Mariooryad, and Carlos Busso. "Audiovisual corpus to analyze whisper speech." *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*. IEEE, 2013.
[24] Saenko, Kate, et al. "Visual speech recognition with loosely synchronized feature streams." *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*. Vol. 2. IEEE, 2005.
[25] Zhao, Guoying, Mark Barnard, and Matti Pietikäinen. "Lipreading with local spatiotemporal descriptors." *Multimedia, IEEE Transactions on* 11.7 (2009): 1254-1265.
[26] Thangthai, Kwanchiva, et al. "Improving Lip-reading Performance for Robust Audiovisual Speech Recognition using DNNs." 2015.
[27] Bear, Helen L., Richard W. Harvey, and Yuxuan Lan. "Finding phonemes: improving machine lip-reading." *FAAVSP 2015* (2015).
[28] Mustafa, Rashed, and Dingju Zhu. "An Efficient Lip-reading Method Using K-nearest Neighbor Algorithm." *TELKOMNIKA Indonesian Journal of Electrical Engineering* 13.1 (2014): 180-186.
[29] Tao, Fei, and Carlos Busso. "Lipreading approach for isolated digits recognition under whisper and neutral speech." *INTERSPEECH*. 2014.
[30] Zhou, Ziheng, Guoying Zhao, and Matti Pietikäinen. "Towards a practical lipreading system." *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. IEEE, 2011.
[31] Shaikh, Ayaz A., et al. "Lip reading using optical flow and support vector machines." *Image and Signal Processing (CISP), 2010 3rd International Congress on*. Vol. 1. IEEE, 2010.

Augmenting ATM Security Analyzing Thermal Imaging and Voice Biometric Recognition

Mehedi Mahmud
CSE Department
University of Asia Pacific
Dhaka, Bangladesh
mehedimahmudcse@yahoo.com

Muhammad F. Mridha
CSE Department
University of Asia Pacific
Dhaka, Bangladesh
firoz@uap-bd.edu

***Abstract*—ATM has become very popular in these days in order to commit teller transaction. The people feel at home using teller machine rather than getting into long line for receiving teller service. This is why this ATM becomes target to the frauds. In this context, biometric security is unparalleled as this security measure deals with individual's biological data. On the contrary, involving biometric features is expensive and that is why the adaptability of this feature is delayed. Thermal image is a single package that informs the system that the subject is alive, provides thermal image to scan and the voice recognition system is used as pass-code security which may provide almost 81% accuracy. Hence, in this paper the researcher intends to accommodate thermal imaging and voice recognition in teller machine as ATM security.**

Keywords—Thermal image, facial recognition, IR, live object detection, voice recognition.

I. Introduction

Financial organizations use several security features in automated teller machines (ATMs) although the crime in this regard has risen to a significant level. The financial organizations like banks, insurances involve security guard and video surveillance and in order to assign such separate security measures these organizations invest huge amount of money and human resources. Nonetheless, the security is outpaced with smart online hackers and schemers. Therefore, it is high time that the helm of the affairs should think of it. The ATM machines should not prone to frauds and should have the ability to outsmart the hackers and schemers. Therefore, a complete package has to be made in ATMs to comply with the total security. This paper analyzes the necessity of using thermal camera incorporated with ATMs which identifies the singularity of the live object, facial recognition and also uses voice recognition as a replacement of ATM PIN (Personal Identification Number). However, complex algorithm is required in order to adopt these facilities in ATMs [1]. Moreover, the accuracy, in this regard, is a great challenge in case of authentication and avoid video spoofing is also a great test for the machine. Therefore, the thermal 3D video gives a significant change in security image as this not only outsmarts video spoofing but also gives firm information regarding the singularity of the person. Moreover, the thermal image gives us the liberty of lighting conditions also. In this paper, another security is augmented and that is the voice recognition as a passcode. Here, the Hidden Marcov Model (HMM) and different search algorithm along with lexical analysis is mediated [2] so that the significance in ATM security rise to a whole new level. The researcher also uses video to string and voice to string algorithm so that the space required for these is lessened.

In this report, the research initially discusses study rationale and latest studies in this context. The researcher also discusses the anti-fraudulent activities. The analysis and experiments are prescribed in the later segment and based on these, the conclusion is drawn.

II. Motivation and Study Rationale

In recent past, the ATM frauds have threatened the management of information security. In ethical aspect, the frauds are so smart that sometimes they are one step ahead. Hence, the motivation of this research is to increase security in cost effective manner. In case of 2D image processing the system cannot differentiate from an image to live object. The color and face symmetry can be understood from the 2D image recognition [3] but this cannot recognize whether the object in front of it, is actual or just a fake image. However, 3D camera can comply with this though the liveliness detection is absent here. Therefore, the thermal imaging is a good solution which also gives solution to determine whether the object in front of it is live or fake. Any material having temperature more than -273.15^o C or 0 K radiate energy as EMP (Electro Magnetic Pulses). Hence, the temperature can be gathered from the thermal image [4] which exhibits different colors to different temperature. Here, emissivity is good measure that how the human body emits energy and from this the psychology of men can also be understood. Moreover, IR measurement is a good choice which annihilates the lighting condition. However, immersion of "Spectra" [5] – a software program lessens the price and this study not only deals with security solution but also obliges cost-effectiveness.

III. Literature Review and Methodology

A. 3D Object Modeling and Recognition

Affine invariant descriptors and spatial relationship between surface patches are the major area of identification in order to recognize the image [6]. In this regard, the geometric representation in different patches is specified to guide the matching process. In case of image, several images are taken from different viewpoints and in case of video moving objects in different dynamic sciences are engaged to vivify the matching. Thus these 3D models give a novel approach for video indexing and matching.

In affine region specification local models of appearance identify the salient image borders and specify the descriptions. Here the surface patches of the same object are projected between images. The adaption shape [7], scale selection [8] as normalized Laplacian, Harris operator [9] to improve position observations are the attributes assigned along with elliptical image border. DoG detector deals with rough surfaces like nose, mouth scaling and surface patches.

(a) Harris-Laplacian b. DoG Detector

Fig. 1. Affine Regions

In normalized representation of image patches and projections represents the affine illumination variations. Here the affine transformation from the rectified specification of image patch vivifies the live scenario. Let R be the affine region of image patch and S be the affine transformation of rectified image patches, here S can be represented as the following 3D matrix:

$$S = \begin{bmatrix} h & v & c \\ 0 & 0 & 1 \end{bmatrix}$$

Here, c is the center of the surface and h and v are the vectors joining c, hence S provides actual form of all representation [10, 11].

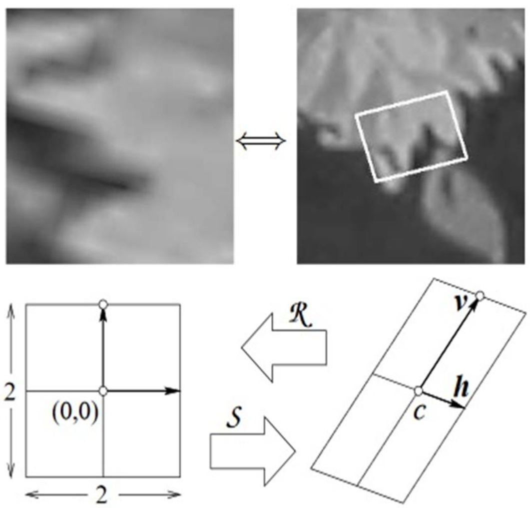


Fig. 2. Geometric shapes

B. *Thermal Image Visualization*

Emissivity specifies the energy emission for an object compared to the emission of energy of black hole. Ambient temperature in IR radiation which implies the temperature of an object and performance, image angle and distance effects are used for thermal image consideration. Every substance radiates energy over 0 kelvin temperature and thus using thermal camera image acquisition is used to perform 3D reconstruction of an object. In case of reconstruction of image triangular mesh from dense point is used. Afterwards, the view is textured and thus the 3D points are defined using intrinsic and extrinsic parameters which specify the polygon [12]. Therefore, total process is as follows:

Image → Point Cloud → Triangulation → Texturing → 3D Model

Fig. 3. Pipeline of thermal image visualization

C. *Speech Recognition*

Hidden Marcov Model is a popular method to implement speech recognition [13]. Time variants under discrete states are identified. The min_HMM (minimal Hidden Marcov Model) in C language recognize speech. The algorithm has two portions, one the front end and other is the search algorithm. In front end there remains an analogue to digital encoder. Afterwards, the digital speech is sampled and analyzed under spectral shaping analysis. The message is then transformed using finite impulse response (FIR) filter. The Signal processing is as follows:

Message → Analog to Digital Encoder → Spectral Shaping → Spectral Analysis → Parametric Transformation → Search Algorithm

Fig. 4. Search Algorithm

The search algorithm uses an observation sequence O and analyze with the preserved word and the mathematical model in this regard as follows:

$$\max p(\omega_i|O),$$

where ω_i is the probability of word and O is the observation vector.

D. *Image & Video Spoofing*

The image and video spoofing techniques gives the liberty to analyze the live human subject. In this case, the blinking based liveliness is analyzed. Here three positions are identified [14]. Visual rhythm analysis is used for video spoofing [15]. General video and image processing is used to convert analogue data to digital format.

IV. Analysis and Work Flow

The idea of this proposed system is to reduce ATM threat which has become a significant fact in case of security. In this report, the authors intend to propose an ATM scenario which identifies the singularity of the user, 3D thermal imaging and uses voice recognition in-lieu-of PIN code. In this regard, the proposed camera is FLIR 757928 one thermal camera which runs in android environment. The most of ATMs run in Windows or UNIX platform. Therefore, the idea is to use Android emulator to run the thermal camera. This camera has good thermal imaging with very affordable price. This shows different color in different temperature. The camera is as follows:

Fig. 4. FLIR 757928 Thermal Imaging Camera

The camera picks images that actually represent human presentation. Therefore the machine can detect that the object standing in front of the machine is actually human. The camera pic representation is as follows:

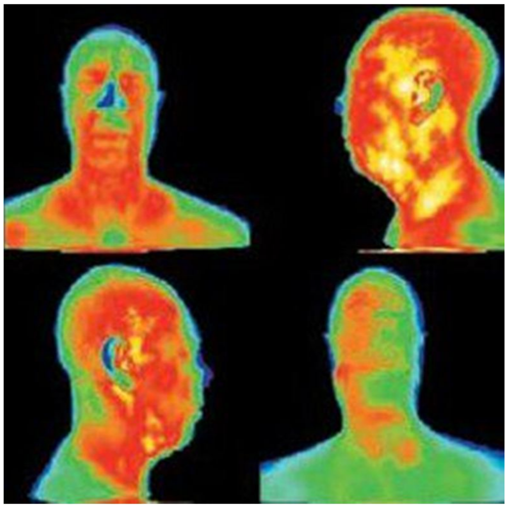


Fig. 5. Thermal Image of Human Object.

The main purpose of using this using this camera is that this not only shows human body presence but also shows total body structure and thus if the user intends to install extra scheming devices, it can be shown in camera due to IR (infrared) camera. Moreover, if someone tries to fake the camera using video or image spoofing technique, it will detect immediately. The color varies in every 0.8°C to 1.2°C. In this regard, the suggestion is that if the camera represents different color, an automated SMS is sent to surveillance and the guard of the ATM. Moreover, the camera also detects all member presence in the room and therefore, if more than one people are present, it is also notified. The user shall use voice passcode instead of PIN here. The voice sample is preserved when registration takes place. The user provides voice passcode which is converted digitally. Then the digital message is sampled and matched using search algorithm. In this regard, the min_HMM is used. Table 1 shows the software and hardware in this regard:

Table 1. Hardware and Software installed

Hardware	Software
FLIR 757928 thermal camera and USB Datacable	Andriod Emulator
Microphone	Analog to Digital Converter – ADCPro
Buzzer/Alarm	Java Simulator

A. *Work Flow Diagram*

The proposed work flow diagram is shown in Fig. 6. In this system, no extra micro controller or accuracy is stored. The matching is conducted in server side. Another java based simulator is installed which send data as soon as the card is inserted in the ATM. All processes are started programmatically. (If data is not matched, buzzer/alarm will ring.) In this regard, it is required to mark in the ground in front of ATM so that the user may understand that he/she has to stand in between the line and the ATM machine. The singularity consideration is a consideration and a definite length is considered for this purpose. The idea is to analyze that the person is alone if someone is waiting behind a queue, he/she needs to maintain a definite distance; otherwise the teller machine will not disburse money and buzzer will ring alarm.

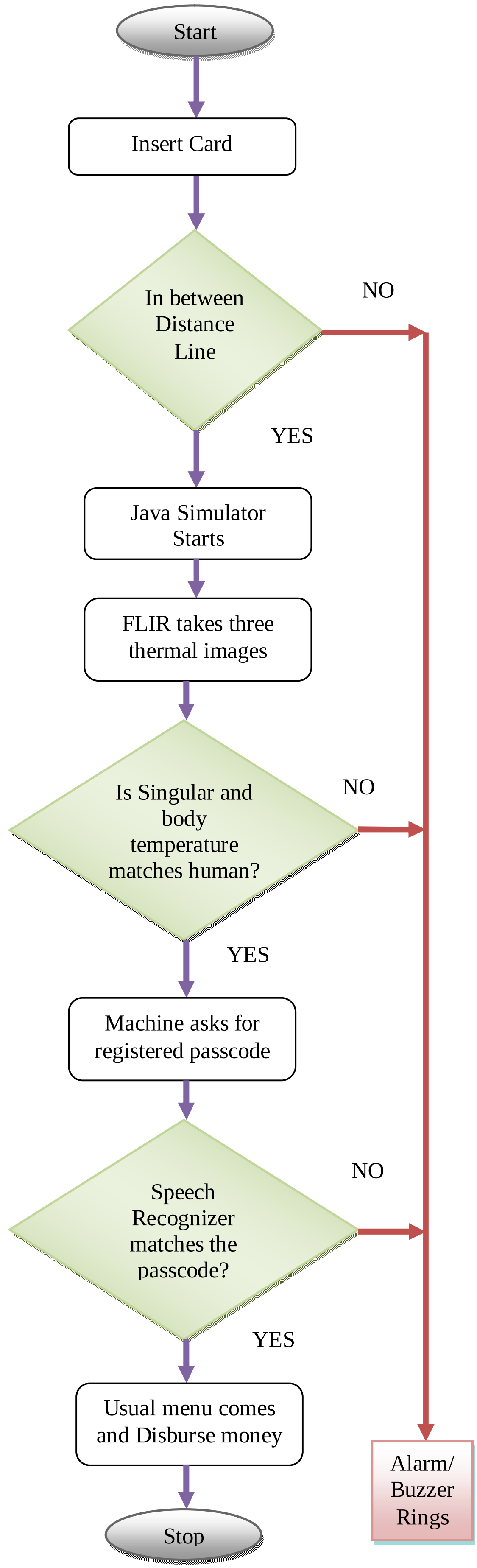


Fig. 6. Flow Chart

V. EXPERIMENTAL EVALUATION

In this section, the authors evaluate the proposed security system. In order to testify the system, the user takes 10 subjects and a laptop where the usual ATM simulator is installed. The purpose of using laptop is that, the ATM machine can be simulated in laptop and in this simulator virtual card can be inserted. The FLIR is installed in the laptop using USB data cable and Java emulator is installed. The laptop is connected to a local LAN network and is able to talk with oracle 11g database where the subjects' registered facial recognition, body temperature, height, voice passcode is stored. The local area is set as 10 GB bandwidth line. The distance in between ATM (actually a laptop which is perceived as ATM) and human subject is 1.5 feet and the distance in between human subject and other humans is set 5 feet. If anyone comes in-between five feet the buzzer rings. In this regard, the noise is a big factor and thus, in case of voice passcode, the band pass filter can be used. In this experiment, it is not used. The microphone of ATM gathers voice passcode. The ADCPro software – analogue to digital converter converts the passcode into voice digital sample. The sample is immediately matched under LAN with the stored passcode. If matches the system gives to do usual operation. The authors also suggest replacing keypad with total voice recognition system so that schemers cannot fake keypad in ATM machine. The FLIR also gives an added facility that if the user contains any unwanted devices, it identifies immediately. In this experiment, six users passed the system, one did not pass due to heavy noise, one used recorded passcode, one used wrong passcode and one was no single. Hence, the system can be tuned in case of noise aspects. The comparison table is given below:

Table 2. Result Table

Users	The-rmal Img	In bet-ween dista-nce line (1.5 Feet)	Live & sing-ular	Facial Recog-nition	Voice Pass-code	Res-ult
User 1	OK	OK	OK	OK	OK	Pass
User 2	OK	OK	OK	OK	OK	Pass
User 3	OK	OK	OK	OK	Heavy Noise	Fail
User 4	OK	OK	OK	OK	OK	Pass
User 5	OK	OK	False	OK	OK	Fail
User 6	OK	OK	OK	OK	OK	Pass
User 7	OK	OK	OK	OK	SNR does not match	Fail
User 8	OK	OK	OK	OK	OK	Pass
User 9	OK	OK	OK	OK	Wro-ng Pass code	Fail
User 10	OK	OK	OK	OK	OK	Pass

Table 3. Users individual attribute

Users	Thermal Image (is taken?) 0/100%	In between distance line (1.5 Feet) 0/100%	Live & singular (Live approach – pass when it is more than 79%)	Facial Recognition (ADCpro Report – pass when it is more than 59%)	Voice Pass code (min_HMM – pass when it is more than 69%)
User 1	100%	100%	91%	67%	91%
User 2	100%	100%	93%	71%	71%
User 3	100%	100%	92%	60%	33%
User 4	100%	100%	99%	66%	73%
User 5	100%	100%	0%	71%	88%
User 6	100%	100%	84%	69%	91%
User 7	100%	100%	89%	61%	5%
User 8	100%	100%	81%	81%	73%
User 9	100%	100%	95%	77%	0%
User 10	100%	100%	83%	68%	76%

The graphical presentation of the attributes of the subjects is given below:

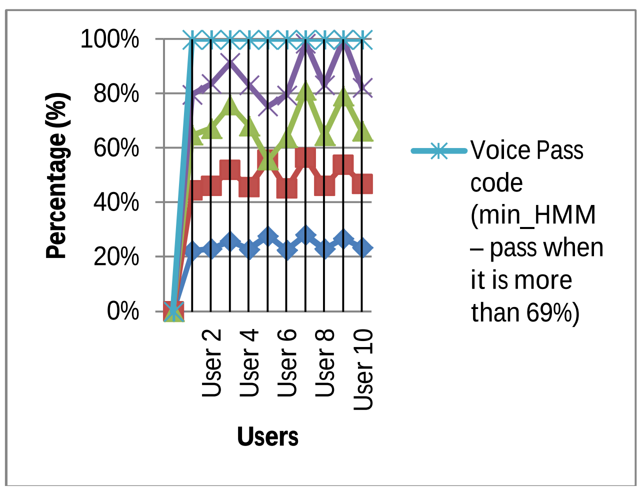


Fig. 7. Linear representation of individual attribute's accuracy of subjects

The accuracy can be presented by the following bar chart:

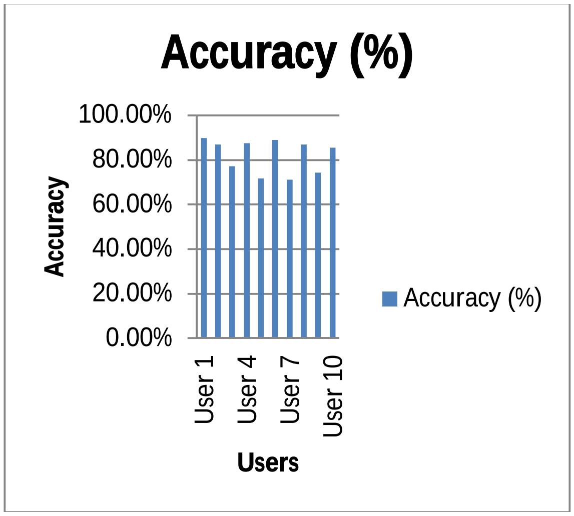


Fig. 8. Bar chart representation of the users' accuracy

The result can be presented as follows:

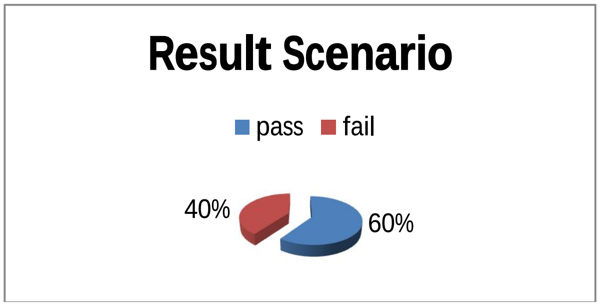


Fig. 9. Result Scenario

VI. Limitations

In this thesis, the researcher has identified five attributes and if any attribute remain absent in the course, the researcher consider this as fail. Therefore, the financial transaction may hamper a bit. Nonetheless, as the time passes, the users will be educated in this regard and this limitation may be dimmed.

The researcher intends to use 3D thermal imaging and thermal video anti-spoofing techniques where the accuracy may rise to 99%; but due to financial cause, the researcher is unable to adopt these in the work.

VII. Future Scope

The Future scope of this work is to adopt 3D thermal imaging and video. The researcher considers various facets in order to be sure that the transaction taking place is absolute. If the researcher includes the 3D thermal imaging and video, the aliveness can be proved easily and thus the absent of one of two attributes can be taken into consideration. Moreover, the idea of using speech recognition may inspire to more artificially intelligent teller machines which may also be able to track of disputes and more denominations of money. Hence, the future directions are positive and this work may lead to encourage artificial intelligent scientists to work in this matter.

VIII. Conclusion

In this work, the ATM is security is enhanced using thermal camera and voice passcode. The authors suggest the voice passcode can be stored in database using Web-RTC and thus, the voice can be stored as string in database. The proposed system not only enhances security, it also enables reliability. The system checks for live object as well as the user's voice passcode. If anyone tries to fake using recorded sound, the signal to noise ratio and the decibel range of his/her voice, will not let it happen. Besides the usual facial recognition is there where the nodal point is set for every user. This proposed work is not only secured but also cost-effective. In the next work more security can be enhanced in order to identify liveliness.

REFERENCES

[1] L. D. inrona and D. Wood, "Picturing Algorithmic Surveillance: The Politics of Facial Recognition Systems", Surveillance & Society CCTV Special (eds. Norris, McCahill and Wood) 2(2/3):177-198, 2004.

[2] L. H. Swee and Texux Instrument, "Implementing Speech Recognition Algorithms on TMS320C2xx Platform", Digital Signal processing solution, 1998.

[3] F. Rothganger, S. Lazebnik, S. Schmidand J. Ponce, "3D Object Modeling and Recognition from Photographs and Image Sequences", University of Illinois, Urbana, USA, 2010.

[4] C. A. Duberstein, S. Matzner, V.I. Cullinan, D. J. Virden, J. Myers and A. R. Maxwell, Automated Thermal Image Processing for Detection and Classification of Birds and Bats. Offshore Wind Technology Assessment. 2012

[5] Spectra. [Online], Available at: http://www.jascoinc.com/spectroscopy/spectroscopy-software, 2015.

[6] A. Aner and J. R. Kender. Video summaries through mosaic-based shot and scene clustering. In European Conference on Computer Vision, pages 388–402, Copenhagen, Denmark, 2002.

[7] J. G˚arding and T. Lindeberg. Direct computation of shape cues using scale-adapted spatial derivative operators. International Journal of Computer Vision, 17(2):163–191, 1996.

[8] T. Lindeberg. Feature detection with automatic scale selection. International Journal of Computer Vision, 30(2):77–116, 1998.

[9] C. Harris and M. Stephens. A combined edge and corner detector. In 4th Alvey Vision Conference, pages 189–192, Manchester, UK, 1988.

[10] D. P. Huttenlocher and S. Ullman. Object recognition using alignment. In International Conference on Computer Vision, pages 102–111, 1987.

[11] D. G. Lowe. The viewpoint consistency constraint. International Journal of Computer Vision, 1(1):57–72, 1987.

[12] J. Wardlaw, M. Gryka, F. Wanner, G. Brostow and J. Kautz. A New Approach to Thermal Imaging Visualization. University College London. 2010

[13] Baum, L.E., and Petrie, T., "Statistical Inference for Probabilistic Functions of Finite-State Markov Chains," Annotated Mathematical Statistics, Volume 37, pp. 1554−1563, 1966

[14] A. Asaduzzaman, A. Mummidi, M. F. Mridha and F. N. Sibai. Improving Facial Recognition Accuracy by Applying Liveness Monitoring Technique. 2016

[15] A.D.S. Pinto, H. Pedrini, W.R. Schwartz, and A, Rocha, "Video-Based Face Spoofing Detection through Visual Rhythm Analysis," IEEE Conference on Graphics, Patterns and Images, pp. 221-228, 2012.

A Survey-based Study on Lip Segmentation Techniques for Lip Reading Applications

Nahid Akhter, Amitabha Chakrabarty

Department of Computer Science and Engineering

BRAC University, 66 Mohakhali, Dhaka-1212, Bangladesh

***Abstract*—** Automatic Speech Recognition plays an important role in human-computer interaction, which can be applied in various vital applications like crime-fighting and helping the hearing-impaired. This paper provides a review of the different technologies used in lip reading and their evolution, from Active Contour Models to Artificial Neural Networks and Hidden Markov Model for temporal Viseme recognition. Special emphasis is placed on techniques in lip-contour detection and tracking, and a comparative study has been presented.

***Keywords*—** *Speech Recognition; human-computer interaction; Viseme , Artificial Neural Networks.*

I. INTRODUCTION

Human-computer interaction is a research area that has fascinated scientists and engineers for a very long time. Within this arena, automatic speech recognition is of special interest as it forms the basis for important human applications, like teaching people with hearing or speech impairment to speak and communicate effectively. Moreover, a visual speech recognition system can help intelligence agencies track a remote conversation by using a camera, where auditory input or support is not available. Visemes,are used by the hearing impaired to view sound visually, thus effectively lip reading the entire human face [7] . Some applications of lip reading include crime fighting potential for computerized lip-reading, speech recognition systems in cars and lip reading systems in computer as an alternative to keyboard. This paper aims to give a run-down of some of the different lip reading techniques, technologies and algorithms that various computer scientists have incorporated thus far, make a comparison among them and present some suggestions.

This paper is organized as follows. First part furnishes the introduction given above. The second part deals with the various steps involved in lip reading. The third part enlists various techniques used in lip reading and their evolution. The fourth section gives an elaborate description of Artificial Neural Networks to be used for learned recognition of syllables. The fifth section provides a brief description of Hidden Markov Models for word recognition and the last section will provide a comparison between various lip contour detection methods.

I. STEPS INVOLVED IN LIP READING

The entire process of lip reading can be broken down into a number of steps. These steps have been listed in chronological order in Fig 1. It starts with Image Acquisition by breaking the video into frames. This is followed by detection of the face, then the lip is detected from the face and the lip ROI is extracted by Lip segmentation. After that, techniques are used to extract important features of the lips and finally Viseme is recognized by a pattern recognition algorithm or technique and the word is recognize by some temporal pattern recognition technique. To detect a human face in a given frame, haarcascade classifiers in OpenCV or Viola Jones algorithm is in MATLAB is used [9].Another algorithm similar to Viola Jones algorithm is the KLT algorithm which can detect and even track tilted and rotated faces, which is not possible in case of Viola–Jones algorithm[10].

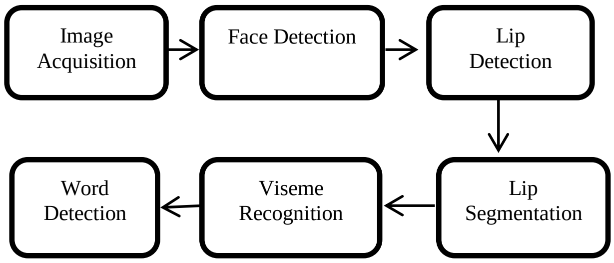


Fig 1. Steps in Lip Reading

For lip detection,Various color spaces such as RGB (Red-Green-Blue), HSI (Hue-Saturation-Intensity), and YCbCr (Luminance-Component blue- Component red) or L*a*b space are used. Pixels of lip area have stronger red component and weaker blue component than other facial regions as proposed in [43]. Therefore, the chrominance component Cr has greater value than the Cb in the lip region. From all the available colour spaces, It has been found in [43] that the Saturation component of the HSI colour space, in combination with the Cr and Cb component of the YCbCr colour space provide a good base .In some cases, the image is converted to a contrast-enhanced black and white image. For better lip recognition, the histeq algorithm is used to enhance contrast so that the lip area appears much darker than the skin.

Lip segmentation techniques may be image-based, colour-based, or model based. The various techniques have been elaborated in Section 6 of this paper. The next step would be extracting the relevant information from the lip segment to be used in recognition of the viseme. Some research papers use pixels of the whole lip image or the just the inner pixels of the mouth as inputs. Other papers advocate the use of certain points on the lip such as the centre of the upper lip, distances between corners of the lip and a few pairs opposite points on the edges of the lip. Whichever technique is used, the end product should be a set of numbers that can be efficiently used as input for a Viseme recognizer.

Invariably and inarguably, the best method for recognizing lip visemes would be modern learning algorithms, called Artificial Neural Networks. These use a set of images to learn standard visemes corresponding to a particular lip image, so that it can recognize a viseme corresponding to a test image. Neural Network model like a Multilayer Perceptron is popularly used for pattern recognition as they can tolerate noise and, if trained properly, will respond correctly for unknown patterns. Sagheer et Al [42] proposed a Hypercolumn Neural Network model to recognize Arabic syllables. Finally, a series of visemes are strung together to recognize a meaningful word, this is done by using Hidden Markov Models.

II. EVELUTION OF LIP READING TECHNIQUES

This section will review some of the popular lip reading techniques available in the literature. This will be just a brief of what each method is capable of doing using what kind of methods.

A. *Snakes*

In 1988, Kass at el [5] proposed the concept of a snake; a model-based technique. Which used an energy-minimizing spline guided by external constraint forces and influenced by image forces that pull it toward features such as lines and edges, as illustrated in Fig2. These snakes lock onto nearby edges, localizing them accurately. Snakes consider different features for image energies like: color of pixels or sharpness of specified area, etc. and provide an account of visual problems like detection of edges, lines and subjective contours; motion tracking and stereo matching. They are guided by user-imposed constraint forces that navigate the snake near and around features of interest. This model is also called an ACM (Active Contour Model).
However, to solve energy minimizing crisis snakes require long computational time and large amount of calculations that make it unfeasible in stand-alone system to extract area function. So, Hashimoto et al [13] introduced variation of the Active Contour Model known as Sampled-ACM. This model assumes area extraction problems as force balancing problems of sample points on the closed curves, which are controlled by three local forces: attraction Fa , Pressure Fp , and repulsion Fr. By calculating the sum of these three forces on each contour point, sampled-ACM can extract the area more rapidly.

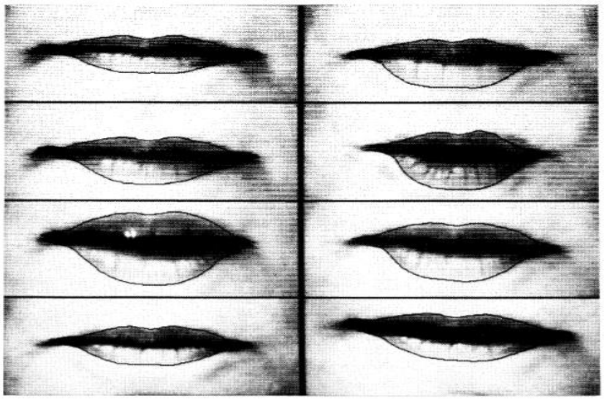


Fig.2.This diagram shows selected frames from a 2-second video sequence using snakes for motion tracking. After being initialized to the speaker's lips in the first frame, the snakes automatically track the lip movements with high accuracy.

Another problem was when the snake is not contacting the object boundary it will enter into the object region and then the repulsion force won't work. So, Sughara at. el. [14] introduced a new force called vibration factor, to improve accuracy against noises in image and combined hardware circuits in FPGA (Field Programmable Gate Array) [12] with the S-ACM vibration factor. This helped improve the area extraction function in standalone systems, but with only one given image.

In 2006, Toshio [15] proposed the Sampled-ACM with splitting characteristics. The proposed Sampled ACM reduced the number of memory accesses required and increased processing speed. Later, using the RASTA-PLP (Relative Spectral Transform - Perceptual Linear Prediction) method a system was built which was robust against any kind of distortions. The pattern recognizer is based on the first n principal components of a 24x16 gray-level matrix centered and scaled around the lips coded to get the outer boundary fairly with "Eigenlips" [16] in regard to the similar approach of Turk and Pentaland's "Eigenfaces" [17], and then it leads on to find the mutual information for audio-visual lip-reading using an MLP (Multi-Layer Perceptron) ANN.

Nowadays, Automatic speech recognition (ASR) is playing a vital role in the human computer interfaces (HCI). Speech recognizers of different kinds were developed in the last decade, like the Tangora system, which is now a STT (Speech to Text) engine [19] [20], created by IBM, which was a large-vocabulary natural-language isolated word recognition system. The first speech recognizer was developed in 1952, which could recognize a single spoken digit by eliminating signal distortion [18]. In 1988, Bahl, at the Watson Research Center described a new type method of word recognition from an acoustic representation of the word [21].

In 1984, Petajan [22] used simple image thresholding to extract binary mouth images, height, perimeter, area and width as visual features to produce their speech reading system. During the last decade, most systems have been based on AV-ASR. In 2000, IBM improved speech recognition by adding visual modality to the traditional audio for Large Vocabulary Continuous Speech Recognition.

B. Neural Networks

Beala and Finaly [23] were the first to apply Neural Networks to the application of lip reading. Rumelhart et al [24], in 1986 used a back propagation system to train a neural network. This went on to be known as a BP network. A BP network can be used to learn and store a great deal of mapping relations of an input-output model. Its learning rule is to adopt the steepest descent method in which the back propagation is used to regulate the weight value and threshold value of the network to achieve the minimum error sum of square. This led to the evolution of systems like NETtalk (Sejnowski and Rosenberg, 1987) for pronunciation of English sentences.

Later, the TDNN (Time-delay Neural Network) network, designed by Stork [25] was introduced that deciphered the time delay and scalability of the system. This network was able to do automatic and acoustic lip reading efficiently in silent as well as acoustic noise environment. Then Yuhas et al. [26], in 1989 accomplished a vowel recognizer using the neural network for static images of mouth shape.

It is noticed that the ANN model developed is accurate enough to recognize the image even if the image is distorted or some portion of data is missing from the image. This model eliminates the long time-consuming process of image recognition.

C. Hidden Markove Model

Later Hidden Markov Models (HMM) were merged with the NN. Further enhancement came about in the form of Modal Neural Network (MNN) [27] for the purpose of robust speech recognition, and Spiking neural network (SNN) designed to cope with neurons that fire multiple spikes in a multi-layer network.

In the classification of dynamic lip movements, geometrical information is extracted from video sequences. The performance of the geometrical-based method remains consistent and unaffected by rotational and brightness changes.

In [45], Yu, Jiang et al. proposed a sentence systematic approach to lip-reading whole sentences by using HMMs integrated with grammar. In this approach, a vocabulary of elementary words is considered. Based on the vocabulary, they define a grammar that generates a set of legal sentences. Each word of the basic vocabulary is modeled by an HMM and the individual HMMs are concatenated according to the rules of grammar. Each HMM corresponding to one of the basic words consists of six states as shown in Fig 3. The HMMs are trained using forward-backward algorithm based on Baum-Welch formula.

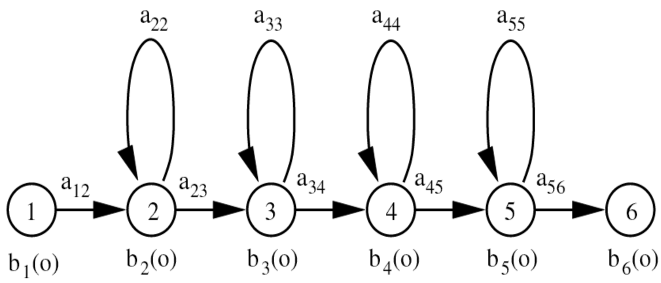


Fig 3. A six-state HMM. Here, O={o_1, o_2,…o_6} is a visual observation of a word, represented by a sequence of feature vectors; A={a_{ij}} is an NxN matrix of state transition probabilities from state i to state j; and B is a set of observation probabilities $b_j(o_t)$ for state j.

Simmons and Cox [46] developed an HMM based system that analyzed a small number of sentences to obtain several acoustic and visual training vectors. Then they created a fully connected 16 state discrete HMM, each state representing a particular vector quantized mouth shape, and producing 64 possible audio codewords. Subsequently, the trained HMM was employed in the Viterbi algorithm to generate the most likely visual state sequence, given the input audio observations.

III. LIP SEGMENTATION TECHNIQUE

For lip reading system, feature extraction is considered as a crucial part. In general, there are two feature extraction methods: 1. “Pixel-based”, 2. “Lip Contour Based”. [2] Table 1 demonstrates the summary of feature extraction/ Lip segmentation methods.

Model	Example	Distortion	Lip Extraction	Performance	Limitation
Image-Based	DCT, DWT, PCA	Real Environ-ment	Visual only	High DCT is better than others	Restricted to illumination, mouth rotation, dimensionality
Model-based	ACM (snakes) , ASM, AAM Deform able templat es	Noise and channel	Acoustic and visual	Low but robust, i.e. invariant to translation, rotation, scaling and illumination	Inner outer lip contours, colour of skin and lip matched, computationally expensive

Image based techniques use the pixel information directly, the advantage is that they are computationally less expensive but are adversely affected by variation such as illumination. Under image-based techniques, there are colour-based and subspace-based techniques. It was found in [30] that the difference between red and green is greater for lips than skin and it was proposed to have a pseudo hue as a ratio of RGB values [31] have also proposed a RGB value ratio based on the observation that blue color plays a subordinate role so suppressing it improves segmentation. Color clustering has also been suggested by some, based on the assumption that there are only two classes i.e. skin and lips. However, if facial hair or teeth are visible, then this does not hold true [6].

In [34] a lip detector based on PCA (A Supspace-based technique) was proposed. Firstly outer lip contours are manually labelled on training data, PCA is then applied to extract the principal modes of contour shape variation, called eigencontour, finally linear regression is applied for detection.

LDA (Linear Discriminant Analysis) has been employed in [35] to separate lip and skin pixels, as shown in Fig 4. [36] have proposed a method in which a Discrete Hartley Transform (DHT) is first applied to enhance contrast between lip and skin, then a multi scale wavelet edge detection is applied on the C3 component of DHT.

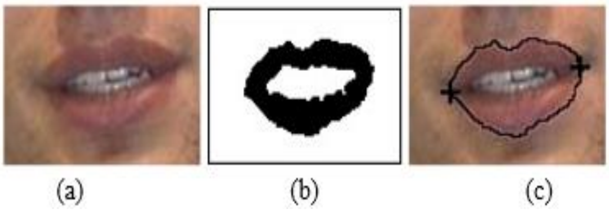


Fig4. (a) The original estimate of the mouth region. (b) Segmented lip region (black) using LDA. (c) The lip contour and the corners of the mouth

Model based techniques are based on prior knowledge of the lip shape and can be quite robust. Snakes have been a commonly used technique for lip segmentation, however, they need to be properly initialized. Moreover, they are unable to detect lip corners as they are located in low gradient regions. So, Eveno et al in [37] proposed a jumping snake which can be initialized far from the lip edge and the parameter adjustment is easy and intuitive[38] proposed a real time tracker that models the dynamic contours of lips using quadratic B-Splines learned from training data using maximum likelihood estimation algorithm. [39] have proposed Active Shape Models (ASM) and Active Appearance Models (AAM), which learn the shape and appearance of lips from training data that has been manually annotated. The introduction of deformable templates led on to proposal of a lip detection method based on Point Distribution Model (PDM) of the face. From [44] it was concluded that the AAM approach produced the most reliable results in terms of lip localization with an error rate of just 0.3%. An example of an AAM fitting has been shown in Fig 5.

In addition, there are some hybrid techniques. These methods combine both image based and model based techniques. Majority of the hybrid techniques proposed in the literature use color based techniques for a quick and rough estimation of the candidate lip regions and then apply a model-based approach to extract accurate lip contours.

Fig 5. An example of AAM fitting. Left column proposes a negative case, right column proposes a positive case.

Usman Saeed and Jean-Luc Dugelay in [6] proposed a “fusion” of edge-based and region-based detection methods to carry out lip segmentation with comparatively better result than any of the two methods carried out individually. Here, given an image, it is assumed that a human face is present and already detected; the first step is to select the mouth Region of Interest (ROI) using the lower one third of the detected face. The next step involves the outer lip contour detection where the same mouth ROI is provided to the edge and region based methods. Finally the results from the two methods are fused to obtain the final outer lip contour. A flowchart of this system is given in Fig 6.

Another method used for lip detection is Fuzzy clustering. This was applied in [32] by combining color information and spatial distance between pixels in an elliptical shape function. [33] have used expectation maximization algorithm for unsupervised clustering of chromatic features for lip detection in normalized RGB color space. Nowadays, lip image segmentation is also done using Fuzzy Clustering as outlined by Leung et al [40]. In this method, multiple clusters are adopted to model the background region sufficiently and a spatial penalty term is introduced to effectively differentiate the non-lip pixels that have similar color features as the lip pixels but located in different regions. Experimental results demonstrate that the proposed algorithm has good segmentation results over other segmentation techniques.

IV. CHALLENGES

Inspite of all these techniques, lip reading comes with its challenges [7]. Firstly, different sounds can be made with the lips in the same position. Secondly, moustaches and beards make lip reading more difficult or even impossible. Thirdly, we form many English sounds in the middle of our mouth, others come from the back of our mouth and even in our throat. These latter are impossible to speech read so far. Moreover, there are numerous homophones in English. Words as different as "queen" and "white" look the same on a person's lips. This accounts to very less accuracy (< 30%) in such a system. Other challenges include fast speech, poor pronunciation, bad lighting, faces turning away, hands over mouths, etc. Most importantly, In order for speech reading to be effective we have to know the subject being discussed.

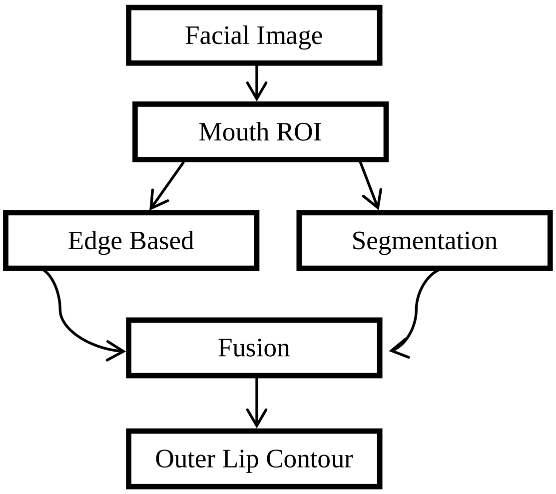


Fig 6. Overview of Saeed and Dugelay's proposed fusion method for lip segmentation

V. CONCLUSION

This paper was created to outline the different research works that have been done in the field of lip reading down the years, while focusing on the various techniques used in lip segmentation, which is a very important step in the process of lip reading. For face and lip localization, Viola Jones algorithm has been unanimously proclaimed as the best way. However, for lip segmentation, there are various methods available, the most primitive of which have been snakes and Active Contour Models. Other techniques include image-based techniques that use differences in colour components of the lips and skin in various colour spaces as well as fuzzy clustering techniques. Techniques that combine both image and model-based methods for lip segmentation were also outlined. Of all the techniques, Artificial Neural Networks for viseme recognition and Hidden Markov Models for Speech Recognition have been found to be the best in modern technologies.

ANNs have been used to detect the viseme being spoken based on advanced learning of previous patterns. Nowadays different researchers are combining different probabilistic, statistic and ANN techniques to provide appreciable and accurate error free automatic Lip-reading systems. HMMs add to the ANNs functionality by providing solutions to the problem of both the acoustic and temporal modeling. Using HMMs, it is now possible to some extent, make predictions of words uttered by evaluating a series of visemes that are stochastically distributed.

Although ANNs try to ape the human central nervous system, and HMMs try to mathematically evaluate real-life phenomena, the performance of current speech recognition systems is far below that of humans. More research needs to be directed to this field as it will have massive impact in real-world applications like helping the deaf understand spoken words crime-fighting institutions pick up clues when auditory input is limited.

REFERENCES

[1] L. R. .Rabiner and B. H. Juang, "An Introduction to Hidden Markov Models", IEEE ASSP Magazine, 1986.

[2] Bushra Naz and Sabit Rahim, "B Audio-Visual Speech Recognition Development Era; From Snakes To Neural Network: A Survey Based Study", Canadian Journal on Artificial Intelligence, Machine Learning and Pattern Recognition Vol. 2, No. 1, 2011.

[3] Toshio Miyaki, Sughara et al, "Active Contour Model with Splitting Characteristics for Multiple Area Extractions and its Hardware Realization", SICE-ICASE International Joint Conference 2006.

[4] Stefan Badura and Michal Mokrys, "Lip detection using projection into subspace and template matching in HSV color space", International Conference TIC, 2012.

[5] Michael Kass et al. "Snakes: Active contour models," International Journal of Computer Vision, pages 321-331, 1987.

[6] Usman Saeed and Jean-Luc Dugelay "Combining Edge Detection and Region Segmentation for Lip Contour Extraction", AMDO'10 Proceedings of the 6th international conference on Articulated motion and deformable objects, Pages 11-20.

[7] Lai Pei Mei. "Interpretation Of Alphabets By Images Of Lips Movement For Native Language", Universiti of Teknologi, Malaysia, 2014.

[8] Shi-Lin Wang et al, "Robust lip region segmentation for lip images with complex background", Science Direct, pages 3481 – 3491, 2007.

[9] "Viola Jones Object Detection Framework", Wikipedia, 2015.

[10] 'Neetu Saini and Hari Singh "Comparison of two different approaches for multiple face detection in color images', International Journal of Innovative Research in Eletrical, Eletronics, Instrumentation and Control Engineering, Vol 3, Issue 1, pages 2321-2004, 2015

[11] Jacek M. Zurada, 'Introduction to Artifiial Neural Systems", 1992 edition, pages 1-21

[12] Md. Khalilur Rahman, "Neural Network using MATLAB" (Powerpoint Presentation), 2005

[13] M.Hashimoto, H.Kinoshita and Y.Sakai, "An Object Extraction Method Using Sampled Active Contour Model," IEICE Trans. D-II, Vol.J77-D-II, No.11, pp.2171-2178, 1994.v

[14] K.Sugahara, T.Shinchi and R.Konishi, "Active Contour Model with Vibration Factor," IEICE Trans. DII, Vol.J80-D-II, No.12, pp.3232-3235, 1997.

[15] T.Miak,T.Kawamura, K.Sughara; "Active Contour Model with Splitting Characteristics for Multiple Area Extractions and its Hardware Realization," SICE-ICASE International Joint Conference, 2006.

[16] C. Bregler and Y. Konig, "Eigenlips for robust speech recognition," in Proc. of ICASSP94, Adelaide, Australia, April 19-22. 1994, pp. 669–672.

[17] M. Turk and A. Pentland, "Eigenfaces for Recognition," Journal of Cognitive Neuroscience, Volume 3, Number 1, MIT 1991.

[18] S.Gurbuz, E.K.Patterson, Z.Tufekci, and J.N.Gowdy Lip-Reading from Parametric Lip Contours for Audio-Visual Speech Recognition;; Department of electrical and computer Engineering Cemson University;Clemson,SC 29634, USA.

[19] 'Pioneering Speech Recognition' from www-03.ibm.com

[20] 'Speech Recognition", Wikipedia, 2015

[21] 'Automatic Speech and Speaker Reognition-Advanced Topics' ,third Edition, Chin-Hui Lee, Frank K. Soong, Kuldip Paliwal, Spinger website, 1999

[22] E. D. Petajan, "Automatic lipreading to enhance speech recognition," in Proc. Global Telecomm. Conf., Atlanta, GA, 1984, pp. 265–272.

[23] Russell Beale , Janet Finlay,"Neural networks and pattern recognition in human-computer interaction," Neural networks and pattern recognition in human-computer interaction, Pages: 460, 1992.

[24] R. P. Lippmann, "Review of Neural Networks for Speech Recognition, Readings in Speech Recognition," A. Waibel and Morgan Kaufmann Publishers, pp. 374-392, 1990.

[25] D.G. Stork, G.WolfP and E. Levinet, "Neural network lipreading system for improved speech recognition," IJCNN, 1992.

[26] B.P.Yuhas, M. H.Goldstein, J.R. and T. J. Sejnowski, "Integration of acoustic and visual speech signals using neural networks," IEEE Communications Magazine, 1989.

[27] H.Kabre, "Robustness of a chaotic modal neural recognition network applied to audio-visual speech," Neural Networks for Signal Processing, page(s): 607 - 616, Sep 1997.

[28] "Inference in Hidden Markov Models", Olivier Cappe, Eric Moulines, Tobias Ryden, Springer, Page 42, 2005

[29] Govind, 'Introduction to Hidden Markov Models' , Lecture 12, CEDAR, Buffalo (Powerpoint Presentation)

[30] Hulbert, A. Poggio, T.: Synthesizing a Color Algorithm from Examples. Science. vol. 239, pp. 482-485 (1998)

[31] Canzlerm, U., Dziurzyk, T.: Extraction of Non Manual Features for Video based Sign Language Recognition. In: Proceedings of IAPR Workshop, pp. 318-321 (2002)

[32] Leung, S.-H., Wang, S-L., Lau, W.-H.: Lip image segmentation using fuzzy clustering incorporating an elliptic shape function. In: IEEE Transactions on Image Processing, vol.13, no.1, pp.51-62 (2004)

[33] Lucey, S., Sridharan, S., Chandran, V.: Adaptive mouth segmentation using chromatic features. In: Pattern Recogn. Lett, vol. 23, pp. 1293-1302 (2002)

[34] Lucey, S., Sridharan, S., Chandran, V.: Initialised eigenlip estimator for fast lip tracking using linear regression. In: Proceedings. 15th International Conference on Pattern Recognition, vol.3, pp.178-181 (2000)

[35] Nefian, A., Liang, L., Pi, X., Xiaoxiang, L., Mao, C., Murphy, K.: A coupled HMM for audio-visual speech recognition. In: Proc. ICASSP, pp. 2013–2016 (2002)

[36] Guan, Y.-P.: Automatic extraction of lips based on multi-scale wavelet edge detection. In: IET Computer Vision, vol.2, no.1, pp.23-33 (2008)

[37] Eveno, N., Caplier, A., Coulon, P.: Accurate and quasi-automatic lip tracking. In: IEEE Transactions on Circuits and Systems for Video Technology, vol. 14, pp. 706 – 715 (2004)

[38] Kaucic, R., Dalton, B., Blake, A.: Real-Time Lip Tracking for Audio-Visual Speech Recognition Applications. In: Proceedings of the 4th European Conference on Computer Vision, vol. II (1996).

[39] Cootes, T. F.: Statistical Models of Appearance for Computer Vision. Technical report, University of Manchester (2004).

[40] Shu-Hung Leung, Shi-Lin Wang, and Wing-Hong Lau, "Lip Image Segmentation Using Fuzzy Clustering Incorporating an Elliptic Shape Function", IEEE Transactions on Image Processing, Vol. 13, No. 1, January 2004.

[41] R. Padilla, C. F. F. Costa Filho and M. G. F. Costa, "Evaluation of Haar Cascade Classifiers Designed for Face Detection", World Academy of Science, Engineering & Technology; 2012, Issue 64, p362.

Comparison of Vocal-Tract Dynamics for Bangla Vowel and Vowel-Consonant-Vowel Sequence

Sathi Rani Mitra
Department of Electrical & Electronic Engineering
Khulna University of Engineering & Technology
Khulna, Bangladesh
Sathi07.eee@gmail.com

Md. Mahbub Hasan
Department of Electrical & Electronic Engineering
Khulna University of Engineering & Technology
Khulna, Bangladesh
hasanatfsg@gmail.com

***Abstract*—In this article, acoustical properties of vocal-tract dynamic-shape for Bangla vowels and vowel-consonant-vowel (VCV) sequences have been analyzed based on dispersion and cross-correlation of linear predictive coding (LPC) filtering coefficients and short-time Fourier transform (STFT) of formant trajectory. Bangla vowels and VCV sequences are recorded from native Bangla speakers and LPC filtering coefficients and formant trajectories are evaluated. The standard deviation of LPC filtering coefficients and transitional-energy of formant trajectories indicate the dynamics of vocal-tract. The consonantal-constriction in VCV sequence accelerates the vocal-tract transitional nature and the transitional nature yields lower-valued cross-correlation with more stable vowels. Numerical values of cross-correlations coefficients are evaluated by LPC trajectories of sounds. Fourier transform technique is utilized to determine the cross-correlation of two unequal length LPC trajectories. These vocal-tract dynamic-behavior representing statistical moments (dispersion, transitional energy and cross-correlation) can be used as consonantal-constriction indication tool in Bangla sounds.**

Keywords—Bangla vowel; Bangla vowel-consonant-vowel (VCV) sequences; vocal-tract; cross-correlation; linear predictive coding (LPC); formant trajectory; short-time Fourier transform (STFT).

I. INTRODUCTION

Acoustical analysis of speech is important for voice activity detection, speech recognition, speech-to-text conversion and many other purposes [1]-[4]. Speech sounds are produced by air pressure vibrations of glottal pulse, which is controlled by the dynamic-shape of vocal-tract and nasal cavities while passing through lips and nasal airways. In modulation perspective, the glottal pulse is modulated by the resonance conditions of the vocal-tract and nasal cavities. Vocal-tract can be considered as flexible tube extends from glottis to lip. During vowel production, the vocal-tract shape remains quasi-stationary [5]. But in the case of vowel-consonant-vowel (VCV) sequence, vocal-tract exhibits transitional shape due to consonantal-constrictions [6]. Serial blending of individual sound related vocal-tract structure does not produce real sounds; it requires fused simultaneous movements of vocal-tract and this process is called co-articulation [7]. Öhman's spectrographic analysis of vowel-consonant-vowel (VCV) sequence utterances results suggested that vowels and consonants are generated by two parallel events and both events are taken place in vocal-tract [8]. The vocal-tract dynamics can be represented by electrical filter which is energized by glottal pulse [9]-[11]. Thus, a VCV is considered to be produced as a vowel-vowel (VV) sequence upon which a consonantal vocal-tract gesture is superimposed.

Linear predictive coding (LPC) filtering coefficients and formant frequency trajectories are usually used to portrait the transfer function of vocal-tract. The LPC or formant trajectories and their dispersions are the members of multi-dimensional acoustical vector which is the mathematical basis of voice recognition [12]-[13]. Euclidean distance between two sound vectors is the separating scale between them. The dispersions of LPC or formant trajectories are prominent members of the multi-dimensional vector. The dispersions of LPC trajectories indicate the dynamics of vocal-tract transfer function. State changing of formant trajectory is another way of presenting vocal-tract dynamics. Formant trajectories are non-stationary types state-variable, and its local spectral energy without dc are the formant transition assisted energy. So the vocal-tract temporal energy of formant trajectory is an identification factor for consonantal attributes in continuous sounds [14]-[15].

The consonantal-transition both in LPC and formant trajectories degrades the similarities between vocal-tract states for vowel and VCV sequence. Cross-correlation between LPC or formant trajectories of two sounds determines the vocal-tract dynamics similarity. The presence of consonantal-constriction in continuous sound can also be identified by the numerical value of cross-correlation with vowels. For this reason, the cross-correlation of sound with vowels are also a consonantal distinguishable factor.

Like any other language, the vocal-tract remains stationary or quasi-stationary for Bangla vowel and moves to transitional state during consonantal-constriction. But the stationary states for vowels and consonantal-transition states are different and these are the prominent factors for voice identification. Recognition cues analysis of Bangla vowels and VCV sequences will be important as Bangla is the most spoken language in Bangladesh and second most spoken language in India, with about 215 million native and about 233 million total speakers worldwide [16]. Research on Bangla speech analysis, recognition and synthesis is in a preliminary stage of development. This preliminary analysis focused on recognition of vowel and different phonemes based on spectral analysis, autocorrelation, LPC coefficients, and mel-Frequency Cepstral Coefficient [17]-[20]. To the best of our knowledge, the numerical investigation of vowel and VCV sequence related vocal-tract dynamics and their comparisons are not reported yet for Bangla. This article makes a comparison of LPC filtering coefficients dispersion and formant transitional energy for Bangla vowels with and without consonantal-constriction. The correlation coefficients are determined for two unequal length LPC trajectories using Fourier coefficient.

This article is prepared as follows: in section 2, the mathematical formulation has been given. In section 3, the process of sounds collection is given. Vocal-tract dynamics related numerical results of experimentally captured sounds are shown in section 4, and at last we will summarize the key points.

II. Mathematical Formulation

A. Acoustic Property of Vocal-Tract

In speech production system, vocal-tract acts as a time-dependent acoustical filter which is excited by glottal pulse and output is speech. The time-dependent filtering characteristics can be expressed by the LPC filtering coefficients. Mathematically, vocal-tract characteristics or transfer function can be expressed by the following equation:

$$H(z) = \frac{S(z)}{U(z)} = \frac{G}{1 - \sum_{k=1}^{p} \alpha_k z^{-k}} \tag{1}$$

where, $S(z)$ is the speech in z domain, $U(z)$ is the glottal pulse in z domain, G is the gain factor, p is the linear prediction order and α_k is the LPC filtering coefficient. The speech samples $s(n)$ can be written by using simple difference equation:

$$s(n) = \sum_{k=1}^{p} \alpha_k s(n-k) + Gu(n) \tag{2}$$

The linear prediction with coefficients α_k and order p is a system whose output is:

$$\bar{s}(n) = \sum_{k=1}^{p} \alpha_k s(n-k) \tag{3}$$

The system function of this linear predictor in z domain is

$$P(z) = \sum_{k=1}^{p} \alpha_k z^{-k} \tag{4}$$

The prediction error is defined as

$$\begin{aligned} e(n) &= s(n) - \bar{s}(n) \\ &= s(n) - \sum_{k=1}^{p} \alpha_k s(n-k) \end{aligned} \tag{5}$$

The error term consists of impulsive glottal pulse with other noises. The coefficients α_k are evaluated by minimizing the predication error $e(n)$. The coefficients α_k are conventionally named as LPC1, LPC2,.....,LPCn [11].

The dispersion (standard deviation μ) of time-dependent LPC filtering coefficients can be determined by the equation:

$$\mu = \sqrt{\frac{\sum_{n=1}^{N} (LPC_n - \overline{LPC})^2}{N}} \tag{6}$$

here, N is the total number of sample, $\overline{LPC}$ is the mean value of samples and LPC_n is the n^{th} value. The dynamic-nature of vocal-tract also changes its resonance frequencies. The resonance frequencies can be estimated from the peak spectral response obtained by using (1). The resonance frequencies are also called formant frequencies.

B. Vocal-Tract Transitional Energy

The changing tendency of formant positions also indicates the vocal-tract dynamics during uttering period. In this article, the vocal-tract transitional energy is formulated by local spectral energies of formant trajectories. As the formant trajectories are non-stationary type, short-time Fourier transform (STFT) is used to obtain transitional-energy. The discrete STFT of the formant trajectory can be expressed as:

$$A_v[n,\Omega] = \sum_{m=1}^{N} F_v[n+m] w[m] e^{-j\Omega m} \tag{7}$$

where, $A_v[n,\Omega]$ is the STFT coefficient and F_v is the v^{th} order formant trajectory. $w[m]$ is the window function. We are interested in transitional energy and $\Omega = 0$ frequency component is neglected here. The total transitional energy of the v^{th} formant trajectory becomes:

$$E_v = \sum_{n=1}^{N} \sum_{\Omega} A_v^{\,2}[n,\Omega] \tag{8}$$

Considering the all formant trajectories, the average transitional energy can be expressed as:

$$E = \frac{1}{N}\sum_{v} E_v \tag{9}$$

C. *Effect of Vocal-Tract Dynamics on Acoustical Correlation*

LPC filtering coefficients refer acoustic property of vocal-tract at the time of sound production and the correlations among the LPC trajectories indicate the similarity among sounds or phonemes. Transitional nature of LPC trajectories are the results of vocal-tract dynamics constriction. The consonantal-transition of vocal-tract will yield lower valued correlation coefficient with vowel. The cross-correlation coefficients between two sounds have been modeled by the summation of zero-lag cross-correlation of LPC trajectory of the two sounds. Mathematically,

$$P_{XY} = \frac{1}{P}\sum_{i}^{P} X_C\left(LPC_{Xi}, LPC_{Yi}, 0\right) \tag{10}$$

Here, P_{XY} is cross-correlation coefficient between X and Y sounds, X_C is the zero-lag cross-correlation coefficient. Generally, the length of LPC_X, and LPC_Y are different, then the zero-lag cross-correlation coefficient has been evaluated using Fourier coefficient and given in (11).

$$X_C = \frac{\sum F_x(w) F_y^{*}(w)}{\sqrt{\sum F_x(w)^2}\sqrt{\sum F_y(w)^2}} \tag{11}$$

here, $F_x(w)$ and $F_y(w)$ are the Fourier coefficients of LPC_X and LPC_Y respectively. The term $\sum F_1(w)F_2^{*}(w)$ is the spectral representation of zero-lag cross-correlation between LPC_X and LPC_Y as:

$$XCorr\ (0) = IFT\left[F_x(w)F_y^{*}(w)\right]$$
$$= \sum_{w} F_x(w)F_y * (w) \tag{12}$$

By using (11), the zero-lag cross correlation between two different length sequence can be computed.

III. Vowels and VCV Sounds Collection

The sounds are collected from male speaker of age 19 to 23 by using Praat software in communication engineering laboratory at the Department of Electrical and Electronic Engineering (EEE) in Khulna University of Engineering & Technology (KUET). Cosonic CT- 863 headphone is used for recording. Eleven Bangla vowels and five VCV sequences are recorded by 5 different male speakers. The speech data was digitized at 44100 Hz sampling frequency and stored as wave format.

IV. Numerically Investigation of Captured Sounds

A. *Acoustic Property of Vocal-Tract*

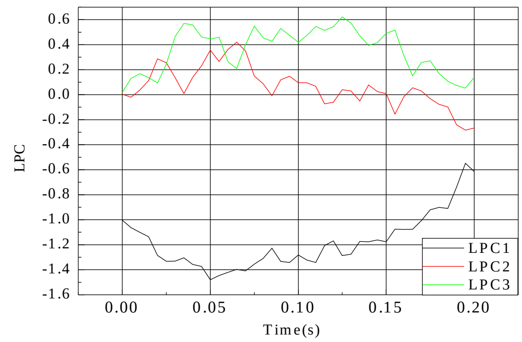


Fig. 1. Evaluated first three LPC trajectories for Bangla vowel.

The 16th order LPC filtering coefficients of Bangla vowels (অ, আ, ই, ঈ, উ, ঊ, ঋ, এ, ঐ, ও, ঔ,) and VCV sequences have been determined every 0.005 second interval using window length 0.025 second. For quick and easy anticipation of 16 LPC trajectories, first three LPC trajectories for vowels (অ) and VCV (ইতি) sequences are shown in Fig. 1 and Fig. 2 respectively. LPC trajectories calculated from experimentally captured speech consists only the vocal-tract states related information not its energized source (glottal pulse). The LPC dispersion related comparisons between vowel and VCV sequence indicates the numerical comparison of vocal-tract dynamics. From visual observation of Fig. 1 and Fig. 2, it is found that more transitions occur during VCV sequence productions. A VCV sequence is formed by a vowel-vowel sequence upon which a consonantal-constriction is superimposed and it can be explained by the LPC trajectories shown in Fig. 2. Higher dispersive LPC trajectories are obtained in the middle due to consonantal-constriction.

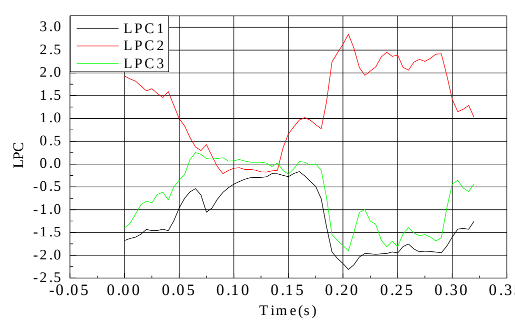


Fig. 2. Evaluated first three LPC trajectories for Bangla VCV sequence.

B. *LPC Dispersion Comparison for Bangla Vowel and VCV Sequence*

TABLE I. STANDARD DEVIATION OF LPC FILTERING COEFFICIENTS OF BANGLA VOWEL

	অ	আ	ই	ঈ	উ	ঊ	ঋ	এ	ঐ	ও	ঔ
LPC1	0.2436	0.29416	0.25384	0.31154	0.21300	0.23060	0.25956	0.27854	0.28692	0.27346	0.36930
LPC2	0.14004	0.14682	0.11708	0.18548	0.16982	0.15868	0.17028	0.16402	0.19730	0.15590	0.23984
LPC3	0.11704	0.14364	0.11524	0.12946	0.09362	0.11032	0.10258	0.14400	0.11670	0.13660	0.14916
LPC4	0.12754	0.14202	0.10158	0.11642	0.09100	0.10204	0.13058	0.12154	0.11308	0.14864	0.12126
LPC5	0.09394	0.10264	0.08170	0.09024	0.09602	0.10122	0.10092	0.10232	0.13368	0.11192	0.11416
LPC6	0.08980	0.10104	0.08930	0.10270	0.07496	0.07312	0.08670	0.10788	0.11258	0.08702	0.08940
LPC7	0.08556	0.11034	0.09280	0.09436	0.07460	0.07304	0.09530	0.13490	0.11290	0.08350	0.09556
LPC8	0.08638	0.10448	0.07926	0.08978	0.06932	0.07472	0.07116	0.08434	0.09076	0.08604	0.08836
LPC9	0.09116	0.09248	0.07250	0.08918	0.07230	0.06978	0.08324	0.08434	0.09756	0.07940	0.08672
LPC10	0.10392	0.10684	0.09716	0.10044	0.09562	0.08188	0.10450	0.10838	0.11594	0.10124	0.12154
LPC11	0.09000	0.09456	0.08300	0.09166	0.08742	0.09436	0.09638	0.10042	0.10686	0.09732	0.09892
LPC12	0.07608	0.08796	0.08596	0.08704	0.08246	0.09392	0.09058	0.09180	0.10662	0.09854	0.09640
LPC13	0.09026	0.09540	0.06314	0.07778	0.07122	0.08082	0.09102	0.08152	0.08594	0.08908	0.09706
LPC14	0.09714	0.09908	0.06830	0.09064	0.06118	0.07184	0.07722	0.08770	0.07530	0.10396	0.08168
LPC15	0.07600	0.09032	0.06578	0.09662	0.06440	0.06850	0.09322	0.08444	0.09850	0.08166	0.07756
LPC16	0.06852	0.07860	0.06944	0.08164	0.0750	0.06744	0.056	0.07588	0.10710	0.06148	0.07242

Dispersion of LPC trajectory for Bangla vowels and 5 VCV sequences (আলো, আমি, ইতি, অনু, উচু) are determined and shown in table I and II, for vowel and VCV sequences respectively. The standard deviations for vowel LPC trajectories are found comparatively small in the range of 0.07 to 0.35. Among the vowels, \অ and \উ have lower-valued and \ঔ has higher-valued standard deviations. From standard-deviation comparisons, it is obvious that \ঔ is the most fricative vowel and \অ and \উ are most sustained vowels among the 11 vowels in Bangla. Vocal-tract dynamic-behavior is the principal factor behind the fricative and sustained nature. For VCV sequences, wide range of standard-deviation (0.10-0.82) is estimated and it indicates the nature of imposed consonantal-constrictions in V-V sequence. Although the consonantal-constriction sustains for a short time, it increases the standard-deviation more than two times than the most fricative vowel. In both vowels and VCV sequences the higher-valued standard-deviations are observed for first five LPC trajectories.

TABLE II. STANDARD-DEVIATION OF LPC FILTERING COEFFICIENTS OF BANGLA VCV SEQUENCE

	আলো	আমি	ইতি	অনু	উচু
LPC1	0.34576	0.28592	0.65646	0.25335	0.59686
LPC2	0.49436	0.52298	0.81716	0.48508	0.63838
LPC3	0.56122	0.52325	0.65832	0.58378	0.61792
LPC4	0.49094	0.50804	0.48288	0.54698	0.44832
LPC5	0.40084	0.40392	0.46924	0.45246	0.31846
LPC6	0.35662	0.42906	0.43928	0.41768	0.30248
LPC7	0.31257	0.40954	0.35538	0.38508	0.27078
LPC8	0.33834	0.36064	0.25634	0.33375	0.29765
LPC9	0.26206	0.29934	0.25472	0.30314	0.23637
LPC10	0.24412	0.29574	0.26866	0.27426	0.27888
LPC11	0.27575	0.28262	0.25376	0.27894	0.24278
LPC12	0.23514	0.28875	0.26622	0.29125	0.24356
LPC13	0.21545	0.29694	0.23842	0.31796	0.21646
LPC14	0.20256	0.27186	0.20904	0.32706	0.19372
LPC15	0.18688	0.22204	0.17172	0.24817	0.15612
LPC16	0.11116	0.13926	0.10856	0.14368	0.10098

C. Transitional Energy of Bangla Vowel and VCV Sequence

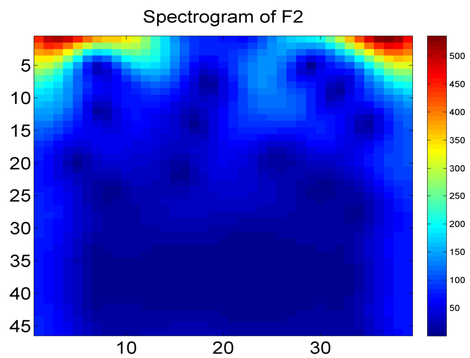


Fig. 3. Spectrogram of 2nd formant trajectory for Bangla vowel.

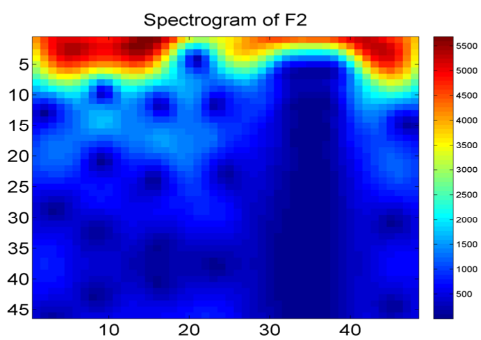


Fig. 4. Spectrogram of 2nd formant trajectory for Bangla VCV sequence.

Formant frequency of the filter defined by LPC trajectories are also evaluated by using Praat. The STFT coefficients of second formant trajectory have been calculated using the formula shown in (7) and the spectrogram for vowel and VCV sequence are shown in Fig. 3 and Fig. 4 respectively. Local frequency components having almost ten times intensity are found in F2 spectrogram of VCV sequence than vowel.

The average transitional energy is calculated by using (9) for vowel and VCV sequence uttered by five different speakers. The average transitional energies for vowel and VCV sequences are tabulated in above table III and IV, respectively. Considering all five speakers, average formant transitional energy is also ten times higher for VCV sequence due to its consonantal-constriction in vocal-tract.

TABLE III. AVERAGE FORMANT TRAJECTORY TRANSITIONAL ENERGY FOR BANGLA VOWEL

Person Name	Average Variation Tendency				
	F1	***F2***	***F3***	***F4***	***F5***
Person 1	2.6E+05	3.14E+05	5.23E+05	5.59E+05	7.60E+05
Person 2	2.2E+05	4.14E+05	1.36E+05	1.11E+05	3.92E+05
Person 3	2.4E+05	3.64E+05	3.29E+05	3.35E+05	5.76E+05
Person 4	2.5E+05	3.29E+05	4.16E+05	4.29E+05	6.50E+05
Person 5	2.4E+05	3.40E+05	5.20E+05	4.20E+05	3.30E+05

TABLE IV. AVERAGE FORMANT TRAJECTORY TRANSITIONAL ENERGY FOR BANGLA VCV SEQUENCE

Person Name	Average Variation Tendency				
	F1	***F2***	***F3***	***F4***	***F5***
Person 1	7.1E+06	7.11E+06	4.38E+06	3.42E+06	5.43E+06
Person 2	1.1E+06	3.40E+06	1.84E+06	2.57E+06	2.56E+06
Person 3	1.9E+06	5.76E+06	5.76E+06	4.19E+06	7.21E+06
Person 4	1.8E+06	4.47E+06	3.12E+06	2.69E+06	4.27E+06
Person 5	1.8E+06	6.42E+06	4.37E+06	3.12E+06	3.12E+06

D. Correlation among Acoustic property of Bangla Vowel Pairs and VCVSequence

To observe the relation among the acoustic property of Bangla vowel, cross-correlation pairs are calculated for each vowel. The correlations of vowels are calculated based upon (10). The correlation coefficient matrix for vowel is shown in table V. From table V, it is found that the value of correlation coefficients among the Bangla vowel is greater than 0.25. So, here is a strong correlation between the acoustic properties of Bangla vowel. Vowels pairs (অ-আ, আ-এ, ই-ঈ, ই-ঋ, ই-ঐ, ঈ-ঋ, ঈ-এ, ঈ-ঐ, উ-ঊ, ঋ-এ, ঋ-ঐ, ঐ-ও) have higher acoustical similarity. The correlations of vowel and VCV sequence are also calculated and shown in table VI. The average value of correlation coefficient for vowel and VCV sequence is 0.17 where the average value for vowel-vowel is 0.55. On average, the consonantal-constriction related vocal-tract dynamics degrade the acoustical similarity 3.24 times.

TABLE V. CORRELATION OF LPC FILTERING COEFFICIENTS OF BANGLA VOWEL PAIRS

	অ	আ	ই	ঈ	উ	ঊ	ঋ	এ	ঐ	ও	ঔ
অ	1										
আ	0.3758	1									
ই	0.3062	0.5406	1								
ঈ	0.4378	0.4574	0.3602	1							
উ	0.5201	0.3103	0.4004	0.4337	1						
ঊ	0.2886	0.3457	0.4673	0.5394	0.6074	1					
ঋ	0.7440	0.3091	0.7337	0.6549	0.3407	0.5588	1				
এ	0.2582	0.3354	0.3385	0.3855	0.4449	0.3557	0.5497	1			
ঐ	0.4018	0.6320	0.7094	0.3832	0.4572	0.3599	0.3376	0.6471	1		
ও	0.5131	0.3447	0.3872	0.4764	0.4976	0.5953	0.3825	0.5016	0.5472	1	
ঔ	0.4295	0.3824	0.4843	0.3975	0.5460	0.4239	0.3559	0.6711	0.6345	0.5187	1

TABLE VI. CORRELATION OF LPC FILTERING COEFFICIENTS OF BANGLA VOWEL AND VCV SEQUENCE

	আলো	আমি	ইতি	অনু	উচু
অ	0.2453	0.2044	0.1238	0.1476	0.2334
আ	0.2463	0.2025	0.1188	0.1610	0.2512
ই	0.1750	0.1684	0.0964	0.1253	0.1822
ঈ	0.1927	0.1878	0.1153	0.1193	0.2217
উ	0.1703	0.1397	0.0780	0.0864	0.1703
ঊ	0.1722	0.1512	0.0950	0.1113	0.1749
ঋ	0.2234	0.2063	0.1373	0.1578	0.2455
এ	0.2532	0.2728	0.1670	0.1841	0.2606
ঐ	0.2136	0.1822	0.1022	0.1614	0.1969
ও	0.2063	0.1752	0.0903	0.1258	0.1682
ঔ	0.1159	0.1098	0.0632	0.0996	0.1070

V. CONCLUSION

Bangla vowels and VCV sequences are analyzed considering the acoustical features related to vocal-tract dynamics during uttering. From LPC filtering coefficients dispersion perspective, \অ and \উ are most stable vowels and \ঔ is the most fricative vowel. Consonantal-constriction on V-V sequence gesture accelerates the transitional state of vocal-tract more than two-time than the most fricative vowel. Formant transitional energy is found more than ten times higher for VCV sequence with respect to vowel. The consonantal-transition of vocal-tract reduces the acoustical correlation with vowel 3.24 times an average.

REFERENCES

[1] S. V. Vaseghi, Multimedia Signal Processing: Theory and Applications in Speech, Music and Communications, John Wiley & Sons, Ltd., 2007.

[2] J. Ramirez, J. C. Segura, C. Benitez, A. de la Torre, and A. Rubio, "Efficient voice activity detection algorithms using long-term speech information," Speech Communication, vol. 42, no. 3-4, pp. 271-287, April 2004.

[3] L. Deng, G. Ramsay, and D. Sun, "Production models as a structural basis for automatic speech recognition," Speech Communication, vol. 22 , no. 2-3, pp. 93-111, August 1997.

[4] R. Reddy and K. S. Rao, "Two-stage intonation modeling using feed forward neural networks for syllable based text-to-speech synthesis," Computer Speech and Language, vol. 27, no. 5, pp. 1105-1126, August 2013.

[5] R. B. Shinde and V. P. Pawar, "Vowel Classification based on LPC and ANN," International Journal of Computer Applications, vol. 50, no. 6, pp. 1-9, July 2012.

[6] J. D. W. Stephens and L. L. Holt, "A standard set of American-English voiced stop-consonant stimuli from morphed natural speech," Speech Communication, vol. 53, no. 6, pp. 877-888, July 2011.

[7] B. H. Story, "Vowel and consonant contributions to vocal tract shape," Journal of Acoustical Society of America, vol. 126, no. 2, pp. 825–836, August 2009.

[8] S. E. G. Öhman, "Coarticulation in VCV utterances: Spectrographic measurements," The Journal of the Acoustical Society of America, vol. 39, no. 1, pp. 151–168, 1966.

[9] H. Wakita, "Direct Estimation of the Vocal Tract Shape by Inverse Filtering of Acoustic Speech Waveforms," IEEE Transactions on Audio and Electroacoustics, vol. AU-21, no. 5, pp. 417-427, October 1973.

[10] B. Yegnanarayana and R. N. J. Veldhuis, "Extraction of Vocal-Tract System Characteristics from Speech Signals," IEEE Transactions on Speech and Audio Processing, vol. 6, no. 4, pp. 313-327, July 1998.

[11] J. D. Markel and A. H. Gray, Jr., Linear Prediction of Speech, Springer-Verlag, 1976.

[12] S. Lopez, P. Riera, M. F. Assaneo, M. Eguia, M. Sigman, and M. A. Trevisan, "Vocal caricatures reveal signatures of speaker identity," Nature Scientific Reports, 3:3407, doi: 10.1038/srep03407, December 2013.

[13] B. S. Atal and S. L. Hanauer, "Speech analysis and synthesis by linear prediction of the speech wave," Journal of Acoustical Society of America, vol. 50, no. 2B, pp. 637-655, August 1971.

[14] P. Niyogi and M. M. Sondhi, "Detecting stop consonants in continuous speech," Journal of the Acoustical Society of America, vol. 100, no. 5, pp. 1063-1076, February 2002.

[15] S. A. Liu, "Landmark detection for distinctive feature-based speech recognition," Journal of the Acoustical Society of America, vol. 100, no. 5, pp. 3417-3430, November 1996.

[16] http:/www.statista.com/statistics/266808/the-most-spoken-language-worldwide.

[17] A. Hasanat, Md. R. Karim, Md. S. Rahman, and Md. Z. Iqbal, "Recognition of Spoken Letters in Bangla," 5th International Conference on Computer and Information Technology (ICCIT), pp. 213-226, 2002.

[18] S. A. Hossain, M. L. Rahman, and F. Ahmed, "Bangla Vowel Characterization based on Analysis by Synthesis," International Journal

of Computer, Information, Systems and Control Engineering, vol. 1, no. 2, pp. 354-357, 2007.

[19] S. A. Hossain, M. L. Rahman, and F. Ahmed, "Acoustic Space of Bangla Vowels," 5th WSEAS Int. Conf. on Signal, Speech and Image Processing, pp. 138-142, Corfu, Greece, August 2005.

[20] A. K. Paul, D. Das, and Md. M. Kamal, "Bangla Speech Recognition System using LPC and ANN," Seventh International Conference on Advances in Pattern Recognition, Kolkata, 2009.

A Holistic Botnet Detection Framework Independent of Botnet Protocols and Architecture

Mohammad Reza Rostami, Norbik Bashah Idris, Zuraini Ismail

Advanced Informatics School, University Technology Malaysia (UTM), Kuala Lumpur, Malaysia
rrmohammad2@live.utm.my, norbik@utm.my, zurainiismail.kl@utm.my

Abstract— **Fast growth of Internet has brought some security concerns. One of these security concerns is Botnet. Bot and Botnets are new sophisticated kind of malware that is equipped with advanced features and have variety of applications. This paper reviews the current botnet detection frameworks and their advantages and drawbacks. To address the drawbacks we propose a conceptual holistic Botnet Detection Framework that is free of limitations regarding the specific botnet protocols and architecture.**

Keywords— Bot; Botnet; Botnet detection framework; network traffic.

I. INTRODUCTION

During the last few years along with the growth of Internet, security concerns have increased as well. Organizations and individuals become more reliant on the Internet services regardless of securely protecting their information. Today, the number of malware threats have increased dramatically due to the financial incentives [1] and rather than simply damaging the client machine, targeted and financially motivated attacks have become further beneficial for the attackers. These types of attacks are harder to detect because they purposely aim to avoid detection, and often are custom-crafted or even encrypted to penetrate signature-based defenses [2]. A new emerged type of malware called Bot [3] includes the aforementioned features as well as advanced capabilities such as controlling and sending commands to the infected computers. Collection of the bot-infected computers creates a malicious network called Botnet [3] and this network receives its updates and commands via a channel from Command and Control (C&C) Servers. The entire network is under control of attacker which is called Bot-Master or Bot-Herder.

Basically, there are three types of botnet Command and Control (C&C) architectures [4] -Centralized, Decentralized and Hybrid- which is the way computers or zombies in the botnet connected to each other and to the servers for receiving new updates. As illustrated in Figure 1, in centralized mode a (or multiple) server(s) controls the entire network and the most used protocols in this architecture are IRC and HTTP. In decentralized mode, as in Figure 2, there are no centralized servers and every zombie is capable to be server and client in this botnet architecture. In other words, this kind of botnet C&C mechanism is based on Peer-to-Peer (P2P) model. The aforementioned architectures have advantages and disadvantages regarding their models and the protocols they use. Therefore, the bot-masters take advantages of each model and create new botnet C&C architecture called Hybrid. For instance, to evade firewall rules the new Waledac (HTTP2P) botnet [4] utilized HTTP protocol over the P2P structure.

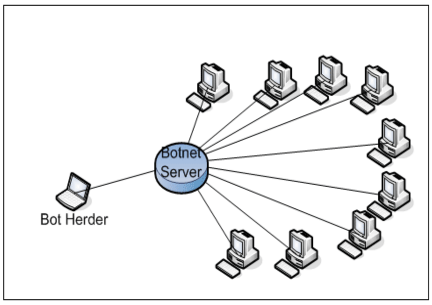


Fig. 1. Centralized C&C

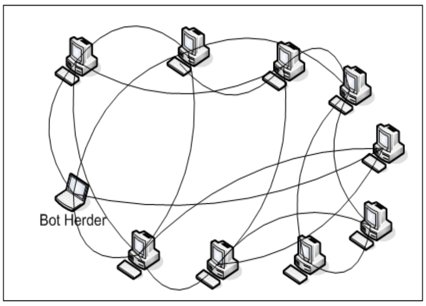


Fig. 2. Decentralized C&C

Currently, several devastating attacks and illegal activities in the Internet are originating from the botnets [5]. Controlling and issuing necessary commands to the infected computers whether to update the bot or performing illicit activities, distinguish and characterize this type of malware compare to the others. To enhance the usage of botnets and obtain much more financial interests, diverse functionalities have been added to the botnets. Thus far, sending SPAM, conducting Distributed Denial of Service (DDoS) attacks and stealing user's confidential information are the main functions of a botnet [6]. However, new features of bots have emerged; for instance, Targeted Attacks to SCADA Systems such as Stuxnet

[7], Duqu [8] and Flamer [9], Search Engine Optimization (SEO) [10] poisonings that divert users to the desired websites (whether it is malicious or just for advertisements).

Botnet detection is integral part of Internet security which needs actions to be taken to have a more secure Internet environment. Thus, several researchers [11-16] have worked on the Botnet Detection. In general, two major approaches exist for detection of the botnets; signature-based and anomaly-based detection. Several researchers [17-19] have studied on the signature-based detection in which their results are applicable for known bots. In this approach, simply every packet is monitored and compared to the pre-configured signatures and attack patterns in the database. Although, in this approach the detection precision is promising, the signature database should be always updated to detect the new bots; otherwise it fails to detect unknown bots. On the other hand, constantly bot-masters obfuscate the bots by novel packers to avoid detection by the signature-based approaches.

Although signature-based detection techniques have remained as a core detection approach, inability to detect unknown bots results in applying anomaly-based detection techniques. Currently majority of researchers [13, 15, 20-22] work on anomaly-based detection approaches to overcome the difficulties of botnet detection. In this approach, based on the rules and heuristics the behavior of network traffic or computer system configurations is classified into normal or anomalous. Anomaly-based detection is divided into host-based detection and network-based detection. Despite the importance of host-based anomaly detection, it is not scalable in the large-scale networks since the monitoring software must be installed in all the computers. Besides, privacy issues have been always a major concern in this method.

In the network-based anomaly detection, the network traffic is analyzed either passive mode or active mode. Several botnet detection frameworks [13, 14, 23] have been proposed based on the network traffic anomaly detection. However regardless of the achieved results, considering limitations for the botnet detection framework, such as detecting specific protocols or pre-defined botnet architecture, make the algorithms deficient regarding the new bots with diverse protocols or architectures. Moreover, the analysis of massive amount of network traffic and encrypted channel between the zombies and the C&C servers are considered as other difficulties of this approach.

Although P2P model is a promising architecture in terms of resiliency [24], the implementation complexities they need to consider and network traffic limitations applied by network administrators, the bot-masters prefer to use HTTP protocol as the main communication medium between the C&C servers and bot infected computers. However, this paper proposes a conceptual botnet detection architecture which is not limited to a specific network protocol. Additionally, because of mentioned limitations for signature-based detection, we use network-based anomaly detection which is capable to even detect unknown botnets.

The rest of the paper is organized as follows. In Section 2, related works and their advantages and drawbacks are discussed. Section 3 states the problems based on the literature review. In sections 4, we propose our conceptual framework along with the system architecture. Conclusion and future works are discussed in section 5.

II. Related Works

Based on the previous section about botnet detection, to collect and understand the behavior of botnets, Honeypot is one of the main techniques of passive monitoring approach. For instance, Andrew [25] used interaction Honeypots to study the attack behaviors of the bot-infected computers.

There are studies worked on specific protocols or pre-defined botnet architecture. Chen [26] examines flow traffic to identify a botnet using HTTP as C&C channel. Houmansadr [27] introduced BotMosaic as a countermeasure to IRC-based botnets. BotMosaic detects particular patterns that are inserted into the network traffic by the watermarker. Liao [28] utilized a methodology based on packet sizes to distinguish P2P botnet traffic from legitimate P2P traffic and the rest of Internet traffic. Choi [29] suggested an anomaly detection mechanism using monitoring group activities in DNS traffic. Based on the group activity model and metric, they developed a botnet detection mechanism, called BotGAD.

Gu et al. suggested vertical and horizontal correlation approaches for botnet detection. BotHunter [17] is based on vertical correlation approach which every host is examined based on its behavior history. For horizontal correlation approach, BotSniffer [14] and BotMiner [13] were developed that examined correlation across multiple hosts. BotSniffer is an anomaly botnet detection system that is able to detect centralized (HTTP and IRC) C&C botnets. Gu [13] claims that BotMiner is not limited to particular protocols or C&C architecture.

The main problem with the previous methodologies is that based on the false positive rate they create, it is ineffective and costly to deploy them in large networks. Thus, Bilge [23], proposed DISCLOSURE that in their framework they use Netflow data for traffic behavior analysis. The major advantage of their framework is that it is applicable in wide-scale dataset. Besides, the researchers apply reputation score to reduce false positive. Their proposed framework can identify C&C servers, which is not applicable for hybrid and P2P botnets. In addition, the features they use can be totally different for various types of botnets and the framework detects 65% of known botnet C&C servers while producing 1% false positive.

Botfinder [12] is another proposed framework that uses high levels of network traffic without deep packet inspection for botnet detection. The researchers use machine learning to identify the key features of C&C communication based on network flow information. This framework also is applicable on a particular bot family that follows certain regular patterns. More important is that it is incapable of detecting botnets engaging in sending SPAM or DDoS attacks. Finally, Botfinder detection rate is 80% that is also varied for different bots. Table 1 summarizes the main botnet detection techniques.

TABLE I. BOTNET DETECTION TECHNIQUES

Botnet Detection Technique	Unknown Botnet Detection	Protocols and Structure Independent
Honeypot [25]	Yes	Yes
Chen [26]	No	No
Liao [28]	No	No
BotGAD [29]	Yes	No
BotHunter [17]	No	Yes
BotSniffer [14]	No	No
BotMiner [13]	Yes	Yes
Disclosure [23]	No	No
Botfinder [12]	No	No

III. PROBLEM STATEMENT

To protect the Internet users, investigation on endpoint protection platforms and signature based software tools continue to block typical botnet threats, but they are incapable and time-consuming on detecting new bots and less effective in detecting low-volume targeted attacks. Development and generation of new bots with remarkable capabilities in incidents of targeted attacks, such as Mariposa [30] and Zeus [31] botnet proves that bot-herders are very well-equipped with in-depth knowledge about security software tools.

The fundamental aspect of botnet detection is to analyze their network traffic behavior in large-scale data to have better conception on the connectivity of bot clients to Command and Control (C&C) servers and vice versa. But administrative restrictions and other technical issues result in ignoring bot analysis in large-scale data.

In addition to ignorance of the analysis part, existing botnet detection approaches are not free of limitations and dependencies regarding the protocols and botnet architecture. As a consequence, to avoid detection bot-masters use multiple communication protocols or infect a computer with other variants of bots differ in structure.

On the other side, to evade detection mechanisms bot-masters try to imitate normal network traffic behavior as well as encrypting the connection channels between bot-clients and C&C servers that result in Deep Packet Inspection (DPI) failure. Accordingly, absence of a generic framework, has led to pose threats into both individual (e.g. losing critical information) and public aspects (e.g. abusing the Internet bandwidth for malicious activities). To address these shortcomings, this paper proposes a conceptual botnet detection architecture which is independent of specific protocols or botnet architecture.

IV. PROPOSED BOTNET DETECTION FRAMEWORK

According to the literature review, existing botnet detection frameworks are lacking of detecting unknown bots. In addition, they are dependent to particular botnet protocol or architecture which makes these mechanisms inefficient. In [32] we analyzed several botnets network traffic in which new network traffic features are introduced.

The proposed technique is utilizing network traffic to detect botnets. Therefore, we are to use network traffic features that separate malicious traffic from benign traffic.

To provide security for the Internet users, it is reasonable to control and analyze the traffic for botnet detection in the edge of the network. Therefore, figure 3 shows that our system is placed in the gateway of the network whether the network is a LAN or an enterprise network (e.g. ISP).

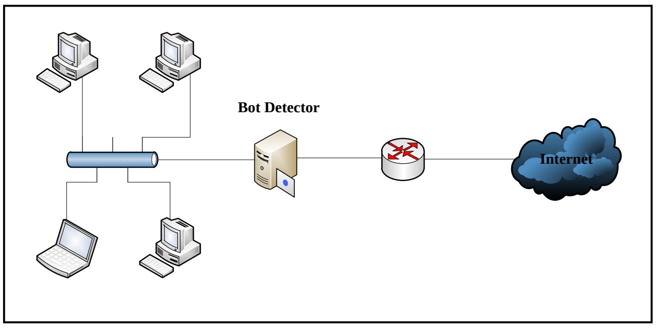


Fig. 3. Proposed Framework

V. CONCEPTUAL SYSTEM ARCHITECTURE

The main part of our proposed framework is Bot Detector. The function of this server is to analyze the network traffic and receive alarms if any botnet evidence has been observed during the analysis of network traffic. The amount of network traffic is very large therefore, therefore to address this problem we propose to use machine learning. While working with machine learning algorithms there should be features (Table 2). With the features these algorithms try to classify the network traffic. As mentioned in previous section the false positive is high because of large network traffic. In order to reduce the false-positive we integrate other security mechanism's logs such as Honeypot, IDS and Firewall.

TABLE II. NETOWORK TRAFFIC FEATURES FOR BOTNET DETECTION

No	Network Traffic Features
1	Duration Mean, Variance, Skewness and Kurtosis
2	Total Bytes Mean, Variance, Skewness and Kurtosis
3	Number of Packets Mean, Variance, Skewness and Kurtosis
4	Average/Variance of inter-arrival time of the packets
5	Number of Failed ICMP Packets
6	Number of TCP Flows
7	Number of UDP Flows
8	Number of SMTP Flows
9	Number of Unique IPs contacted
10	Number of suspicions Ports
11	Request frequencies
12	Requests intervals
13	DNS- Number of concurrent queries
14	DNS- Cumulated queries to each distinct domain

15	DNS- Query frequency
16	Duration Mean, Variance, Skewness and Kurtosis

Figure 4 shows the architecture of our proposed botnet detection system which consists of 4 phases: Initialization Phase (IP), Analysis and Inspection Phase (AIP), Detection Phase (DP) and Report Phase (RP).

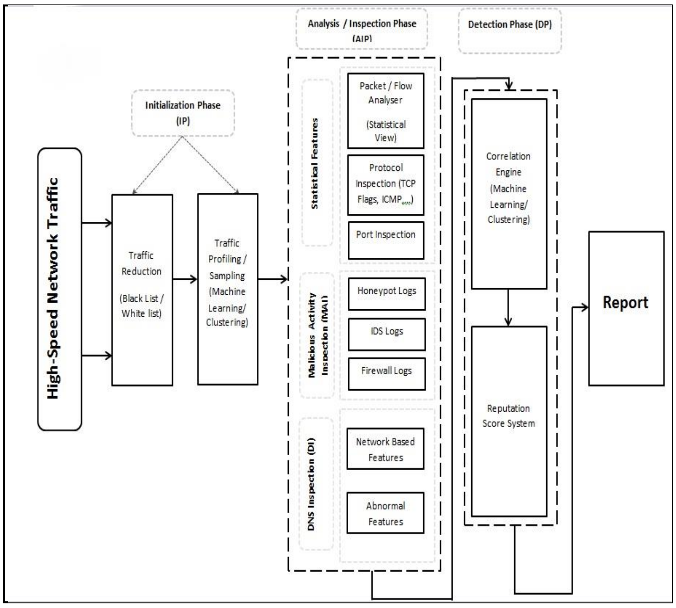


Fig. 4. Conceptual System Architecture

The proposed system is deployed at the edge of network and it works in passive mode. Each phase has some tasks to do.

• Initialization Phase (IP): To prepare the data for better analysis, some pre-processing actions are conducted on the network traffic such as traffic reduction based on Black/White list and traffic profiling.

Certainly for IP or domain white listing, we only consider the IP address of the main page of top most popular websites like Google or Microsoft. Because other services that are offered by these companies can be used for malicious activities. For instance, Gmail or Google Drive can be used by bot-master to deliver the latest updates to bots.

In the traffic profiling, a history of each computer is preserved for further analysis. All the visited domain names, destination IP addresses, request frequencies and intervals that have been made to destination addresses or websites, number of concurrent DNS resolutions, accessing to an IP with unusual port numbers, will create a behavioral profile for each computer in our network. This profile are used for further analysis in the next phases.

• Analysis and Inspection Phase (AIP): This phase analyses the required factors that is helpful for botnet detection. Along with the packet features such number of packet per second, total bytes per second, flow features also are utilized. The flow features are flow duration, number of packets per flow, number of TCP flows, number of UDP flows and number of SMTP flows. In our botnet analysis [33], there are several failed ICMP packets. In P2P botnets, bots try to communicate with the IP addresses that are hard-coded in their peer lists. After some time, the infected computers could be disconnected from network or the bot is detected by the antivirus. Thus, there are many failed ICMP packets when analyzing P2P botnet network traffic. This can be a valuable feature for P2P botnet detection.

Spam bots joins the infected computers to their email spam campaigns. Therefore, while investigating the network traffic, numerous SMTP packets exist which is very unusual for a computer that is not Email server. Accordingly, during identification and tracking of botnets it is very essential to consider various related protocols.

Apart from the mentioned features, to decrease the false positive and achieve higher detection accuracy rate Honeypot, Intrusion Detection System (IDS) and firewall logs are also considered as extra features. Essentially, these network security mechanisms work based on signatures which can be very valuable while utilizing machine learning classifiers. The generated logs are used as features to ascertain the identified malicious traffic.

When a computer is infected with a bot, initially bots try to stablish a connection with C&C servers to gain commands or updates. By performing DNS queries the bots locate the particular server. Hence, it is possible to use DNS traffic statistics to create a detection mechanism which looks for DNS anomalies.

Examining DNS queries, especially those consecutive queries might provide significant information about botnet existence and C&C server locations. Other important features of DNS queries can also be useful for botnet detection, including query frequency, query lifetime, domain lifetime.

• Detection Phase (DP): The main phase of this architecture that receive required statistics from the previous phase and cluster the IP addresses based on their features mentioned above. Experimentally, bots within the same botnets present similar network traffic patterns as they are programmed to perform regular communication with the C&C server. Moreover, they use same protocol when they tend to receive updates or any command from the bot-master. Therefore, based on the received statistics from analysis and inspection (AIP) phase and mentioned features, similarities among the network patterns are extracted.

For well-known botnets, machine learning classifier is utilized. The classifier is trained with various network traffic data and is able to detect famous botnets accurately. To detect unknown botnets, the results of classifier which are labeled as benign traffic are fed to hierarchical clustering algorithm which builds different clusters based on distance connectivity. In each clusters, there are several network flows with similar network traffic features such as destination IP address, port numbers, number of packets in a flow and etc. The output of this phase is sent to security expert for further analysis and if it is found out that some clusters are consisting botnet traffic then these clusters are used to train our machine learning classifier with new botnet network traffic features.

The output of this part goes into the reputation score system and compare with the previously detected known IP addresses.

- Report: The final phase reports the results of analysis to the network administrator for further analysis in which IP address can be suspicious or normal.

VI. Conclusion and Future Work

Botnets and their applications are growing very fast and botnet detection regardless of the protocols and architecture is a challenging problem. In this paper, we proposed our conceptual holistic botnet detection framework. Our system is free of any limitations related to botnet protocols and architecture. To reduce the false positive, we integrate other security mechanisms such as Honeypot, IDS and Firewall Logs.

In our future work we will explain our framework in details with the achieved results along with applying machine learning algorithms and probabilistic theories.

Acknowledgement

The work is funded by UTM GUP Research Grant #PY/2015/05229.

References

[1] L. Orans and P. Firstbrook, "Magic Quadrant for Secure Web Gateways," *Gartner, May,* vol. 25, 2011.

[2] S. S. Silva, R. M. Silva, R. C. Pinto, and R. M. Salles, "Botnets: A survey," *Computer Networks,* vol. 57, pp. 378-403, 2013.

[3] J. Kwon, J. Lee, H. Lee, and A. Perrig, "PsyBoG: A Scalable Botnet Detection Method for Large-scale DNS Traffic," *Computer Networks,* 2016.

[4] M. Eslahi, R. Salleh, and N. B. Anuar, "Bots and botnets: An overview of characteristics, detection and challenges," in *Control System, Computing and Engineering (ICCSCE), 2012 IEEE International Conference on*, 2012, pp. 349-354.

[5] M. Grottke, A. Avritzer, D. S. Menasché, L. P. de Aguiar, and E. Altman, "On the Efficiency of Sampling and Countermeasures to Critical-Infrastructure-Targeted Malware Campaigns," *ACM SIGMETRICS Performance Evaluation Review,* vol. 43, pp. 33-42, 2016.

[6] W. Yong, S. H. Tefera, and Y. K. Beshah, "Understanding Botnet: From Mathematical Modelling to Integrated Detection and Mitigation Framework," in *Software Engineering, Artificial Intelligence, Networking and Parallel & Distributed Computing (SNPD), 2012 13th ACIS International Conference on*, 2012, pp. 63-70.

[7] R. Langner, "Stuxnet: Dissecting a cyberwarfare weapon," *Security & Privacy, IEEE,* vol. 9, pp. 49-51, 2011.

[8] J. P. Chapman, E. Gerhards-Padilla, and F. Govaers, "Network traffic characteristics for detecting future botnets," in *Communications and Information Systems Conference (MCC), 2012 Military*, 2012, pp. 1-10.

[9] M. S. Vinay and M. Balakrishnan, "A Comparison of Three Sophisticated Cyber Weapons," *Managing Trust in Cyberspace,* p. 387, 2013.

[10] D. Y. Wang, S. Savage, and G. M. Voelker, "Juice: A Longitudinal Study of an SEO Botnet," in *NDSS*, 2013.

[11] L. Sheng, L. Zhiming, H. Jin, D. Gaoming, and H. Wen, "A Distributed Botnet Detecting Approach Based on Traffic Flow Analysis," in *Instrumentation, Measurement, Computer, Communication and Control (IMCCC), 2012 Second International Conference on*, 2012, pp. 124-128.

[12] F. Tegeler, X. Fu, G. Vigna, and C. Kruegel, "Botfinder: Finding bots in network traffic without deep packet inspection," in *Proceedings of the 8th international conference on Emerging networking experiments and technologies*, 2012, pp. 349-360.

[13] G. Gu, R. Perdisci, J. Zhang, and W. Lee, "BotMiner: Clustering Analysis of Network Traffic for Protocol-and Structure-Independent Botnet Detection," in *USENIX Security Symposium*, 2008, pp. 139-154.

[14] G. Gu, J. Zhang, and W. Lee, "BotSniffer: Detecting botnet command and control channels in network traffic," 2008.

[15] M. Stevanovic and J. M. Pedersen, "An analysis of network traffic classification for botnet detection," in *Cyber Situational Awareness, Data Analytics and Assessment (CyberSA), 2015 International Conference on*, 2015, pp. 1-8.

[16] S.-H. Li, Y.-C. Kao, Z.-C. Zhang, Y.-P. Chuang, and D. C. Yen, "A Network Behavior-Based Botnet Detection Mechanism Using PSO and K-means," *ACM Transactions on Management Information Systems (TMIS),* vol. 6, p. 3, 2015.

[17] G. Gu, P. Porras, V. Yegneswaran, M. Fong, and W. Lee, "BotHunter: detecting malware infection through IDS-driven dialog correlation," presented at the Proceedings of 16th USENIX Security Symposium on USENIX Security Symposium, Boston, MA, 2007.

[18] J. Goebel and T. Holz, "Rishi: Identify Bot Contaminated Hosts by IRC Nickname Evaluation," *HotBots,* vol. 7, pp. 8-8, 2007.

[19] J. Zhang, R. Perdisci, W. Lee, U. Sarfraz, and X. Luo, "Detecting stealthy P2P botnets using statistical traffic fingerprints," in *Dependable Systems & Networks (DSN), 2011 IEEE/IFIP 41st International Conference on*, 2011, pp. 121-132.

[20] O. Pomorova, O. Savenko, S. Lysenko, A. Kryshchuk, and K. Bobrovnikova, "A Technique for the Botnet Detection Based on DNS-Traffic Analysis," in *Computer Networks*, ed: Springer, 2015, pp. 127-138.

[21] D. Zhao, I. Traore, B. Sayed, W. Lu, S. Saad, A. Ghorbani*, et al.*, "Botnet detection based on traffic behavior analysis and flow intervals," *Computers & Security,* vol. 39, pp. 2-16, 2013.

[22] J. Zhang, R. Perdisci, W. Lee, X. Luo, and U. Sarfraz, "Building a scalable system for stealthy p2p-botnet detection," *Information Forensics and Security, IEEE Transactions on,* vol. 9, pp. 27-38, 2014.

[23] L. Bilge, D. Balzarotti, W. Robertson, E. Kirda, and C. Kruegel, "Disclosure: detecting botnet command and control servers through large-scale netflow analysis," in *Proceedings of the 28th Annual Computer Security Applications Conference*, 2012, pp. 129-138.

[24] D. Andriesse, C. Rossow, B. Stone-Gross, D. Plohmann, and H. Bos, "Highly resilient peer-to-peer botnets are here: An analysis of Gameover Zeus," in *Malicious and Unwanted Software:" The Americas"(MALWARE), 2013 8th International Conference on*, 2013, pp. 116-123.

[25] D. Andrew and H. Chi, "An empirical study of botnets on university networks using low-interaction honeypots," in *Proceedings of the 51st ACM Southeast Conference*, 2013, p. 44.

[26] C.-M. Chen, M.-Z. Huang, and Y.-H. Ou, "Detecting Web-Based Botnets with Fast-Flux Domains," in *Advances in Intelligent Systems and Applications-Volume 2*, ed: Springer, 2013, pp. 79-89.

[27] A. Houmansadr and N. Borisov, "BotMosaic: Collaborative network watermark for the detection of IRC-based botnets," *Journal of Systems and Software,* vol. 86, pp. 707-715, 2013.

[28] W.-H. Liao and C.-C. Chang, "Peer to peer botnet detection using data mining scheme," in *Internet Technology and Applications, 2010 International Conference on*, 2010, pp. 1-4.

[29] H. Choi, H. Lee, and H. Kim, "BotGAD: detecting botnets by capturing group activities in network traffic," in *Proceedings of the Fourth International ICST Conference on COMmunication System softWAre and middlewaRE*, 2009, p. 2.

[30] P. Sinha, A. Boukhtouta, V. H. Belarde, and M. Debbabi, "Insights from the Analysis of the Mariposa Botnet," in *Risks and Security of Internet*

and Systems (CRiSIS), 2010 Fifth International Conference on, 2010, pp. 1-9.

[31] L. M. Ibrahim and K. H. Thanon, "Analysis and Detection of the Zeus Botnet Crimeware," *International Journal of Computer Science and Information Security,* vol. 13, p. 121, 2015.

[32] M. R. Rostami, M. Eslahi, B. Shanmugam, and Z. Ismail, "Botnet evolution: Network traffic indicators," in *Biometrics and Security Technologies (ISBAST), 2014 International Symposium on*, 2014, pp. 274-279.

[33] M. R. Rostami, B. Shanmugam, and N. B. Idris, "Analysis and detection of P2P Botnet connections based on node behaviour," in *Information and Communication Technologies (WICT), 2011 World Congress on*, 2011, pp. 928-933.

Performance Evaluation of VoIP Services over UMTS-Network with Differentiated UMTS Bearer Services

Dhawa Sang Dong, Anand Gachhadar
Department of Electrical and Electronics Engineering
School of Engineering, Kathmandu University, Dhulikhel, Nepal
Email: dhawa1984sdong@gmail.com, anand.gachhadar@ku.edu.np

***Abstract*-- Quality of service (QoS) for voice over IP network is not consistent because of absence of dedicated links as in circuit switched (CS) network. Because of wider mobility to users in UMTS network and voice over IP network being low cost service, performance evaluation for VoIP service over UMTS network is realized more important. Prioritized forwarding strategy to different traffics can be controlled allocating Forward Access Channel (FACH) Scheduling weights in Radio Network Controller (RNC) and Resource Reservation Protocol (RSVP) for the resource allocation like buffer space, bandwidth etc.**

In this paper, End-to-End QoS of UMTS is mapped to external IP network with Differentiated Service Code Point (DSCP) field in IP packet header specifying different per hop behavior (PHB) and significant improvement in QoS of VoIP service over UMTS network has been observed. Hybrid model of QoS enhancement with DiffServ Code Point integrated to FACH scheduling weight inside UMTS network and RSVP scenario has been studied in this work and the model has been purposed to reduce End-to-End delay maintaining better Mean Opinion Score (MOS).

Keywords: QoS, DSCP, FACH Scheduling, RAB service, CNB service.

I. Introduction

As the circuit switched (CS) network is specially designed for voice signal transmission, the network has good qualitative speech in terms of MOS value and good spectral efficiency in terms of voice band [2]. Gateway GPRS service node (GGSN) enables access to external IP network without multimedia that is data services only while IP Multimedia Subsystem (IMS) defined 3GPP releases supports multimedia services too [13]. QoS is an End-to-End measure which is to be met through all the modules in the network where different applications have its own different QoS requirements [12]. When the voice packet are transmitted over packet switched (PS) network, the quality of services of the network must have closer performance in terms of QoS parameters that CS network can provide [2] and the main purpose of QoS architecture [10] in UMTS network is to optimize the delivery services to preferential service quality to the application with optimal data rate, reducing data loss and controlling latency. QoS architecture defined by 3GPP consists of UMTS Bearer services consisting Core Network Bearer (CNB) services and Radio Access Bearer (RAB) services [9] with RAB service ensuring transmission of control signal and user traffics between CN and MT and CNB efficiently controls the UMTS Backbone Network utility to ensure the UMTS QoS. With the increased data speed and high mobility services in PS networks, data over internet is becoming better choice for users in terms of cost and better option for service provider in terms of resource management and so the voice application is “de facto standard” in mobile terminals [3].

Depending up on the quality of signaling at the time of call setup and call termination, delivery quality and call quality, voice communication over any network is categorized as good quality or not. Performance of a VoIP service can be analyzed with the parameters End-to-End delay (including queuing delay), packet loss, jitter, throughputs and traffics intensity over the network. E-model as specified by the standard ITU-T G.107 gives the impairment factor (R) ranging from 0 to 100 with linear relation with perceived quality of voice [2]. This R factor can be transformed to MOS scale value with 3.5 or above to be acceptable quality of voice [2].

In 2G and 2.5 G wireless network, voice signal is transmitted over CS network while data signals are switched over GPRS network or infrastructure [4]. Different signaling protocols such as session initiation protocol (SIP) or H.323 are used to establish session between end users with appropriate retransmission timer [4]. SIP can use either TCP or UDP transport protocols while H.323 v1/v2 can use TCP protocols for transmission of packets [4]. Optimum selection of signaling protocols and transport protocol determines the call set up performance. In [4] different combination of protocol stacks, SIP, UDP and RLP, SIP, TCP and RLP, H.323 and

TCP, are studied and SIP/UDP/RLP with adaptive retransmission timer was found to be optimum.

Since voice signal is analog, codec GSM-FR or G.729A is used to encode the voice stream at transmitting end, i.e. digitization of voice stream to transmit over IP network [1]. Encoded signal or bit streams are to be decoded to analog signal at the receiving end which is performed by codec [1].
Transmission rate can be controlled with Decision Feedback Scheme with RTCP(Real Time Transmission Control Protocol) where RTCP gives information about network conditions and quality level of reception, consequently congestion is controlled [5]. Rate control unit adopt the transmission rate according to round trip time and End-to-End packet loss [5]. Two bytes of field from error detection and correction control field in UDP header of VoIP are removal which is supported by 3GPP [6] reducing payload in UMTS network and power consumption.

A hybrid model to improve QoS is proposed in this study for VoIP service over UMTS network. In the model proposed, G.729A codec with 1ms frame size is used with DiffServ Code Points Expedited Forwarding (EF) to conversational type traffic, Assured Forwarding (AF) to streaming and interactive type traffics and Best Effort (BE) to background traffic and with RSVP and FACH scheduling where FACH is used to transport control signal and some user data in transport channel to secondary common control physical channel in physical layer whereas DCH is used for user data with dedicated channel in transport layer and mapped to dedicated physical data channel in physical layer [11].

Rest of the paper is structured as follow. Section II deals with literature review, section III with network simulation model in OPNET, section IV with evaluation of results from simulation and analysis and section V concludes the whole work and studies done in this paper.

II. Literature review

In [4], session set up time attention has been paid, is affected by FER which is measure of quality of wireless link. H. Fathi and et al. in [4], has evaluated setup time for VoIP session within the range of 0-10 percent of FER at radio link showing significant reduction with Radio Link Protocol (RLP) and average of 46% decrement was found with adaptive transmission timer. Signaling protocol H.323 and SIP have comparable performance for FER less than 2% while greater than 2% SIP has better performance compared to H.323.

At primitive stage of VoIP technology, Cuny and Lakaniemi in 2003 [7] evaluated end-to-end delay of 221.96 ms ranging FER 0 to 5% at radio link is developed to end-to-end delay of 150 ms with frame size of 20 ms in 2008 [1]. E.R. Vale, et al got 6 bytes of reduction to high probability signal level in VoIP is obtained with roughly 15% of power saving [6] if the network code of error correction and detection is efficiently developed.

Dimension of de-jitter buffer size and Block Error Rate(BLER) has significant impact on QoS of the PS network and BLER of 2% has transmission delay of 132.7 ms which is less than ITU-T G.114 specification [2]. MOS of value 3.91 was evaluated as acceptable value greater than threshold of value 3.5 at BLER 2% and buffer size of 180 ms and maximum playout interruption at zero buffer size and 19.2 at buffer size of 180 ms at BLER 2% [2]. Sivabalakrishnan and Manjula concerned about the improvement of QoS with adaptive data transmission rate with the network congestion level implementing Decision Feedback Scheme (DFS) [4], using RTCP which uses senders' report and receivers' report to estimate the data rate consequently packet loss and delay can be reduced focusing on optimum utilization of bandwidth available.

20 ms frame size was found to be better option for the codec G.729A and GSM-FR while for G.711 high packet losses was found in terms of end to end delay which has direct impact on MOS for voice quality [1]. Frame numbers per VoIP packet has effects on QoS as increased number results increasing delay a frame of 20 ms per VoIP packet was found to be best for GSM-FR and G.729A Codec [1]. Mallah and Islam [10] studied the impacts of data rate on QoS for different data traffics classes streaming, background, interactive and conversational class type and studied how Forward Access Channel scheduling affects the traffics over UMTS network. In [10], Mallah and Islam analyzed the mean end-to-end delay for data rate of 60 kbps and for different combination of scheduling weight to different traffic classes and found scheduling has significant impact on end-to-end delay.

Shreevastav and et al. proposed two layer scheduler (TLS) to improve QoS and cell throughput [8] and in the proposed algorithm, first layer performs resource reservation with the weight assigned based on delay sensitivity. Quality of radio link periodically measured by UE is reported as channel quality indicator (CQI) to corresponding NodeB and is used to ensure the packet delivery in the algorithm [8] where users with higher buffer occupancy and enough CQI value is dequeues first among the other users by scheduler. The simulation in NS-2 environment for proposed algorithm shows better throughput among other algorithm with improved jitter and delay in [8] indicated with peak signal to noise ratio (PSNR) results. A similar analysis was performed by Saad and et al in [12] that isstudied end to end QoS for DiffServ IP Network model and compared the result to their hybrid model DiffServ/RSVP concluding significant improvement in End-to-End delay, packet loss, delay variation and throughput.

Jadhav and et al in [3] investigated the end-to-end delay, MOS, jitter and delay variation with the discovery of WiMAX with outstanding performance over UMTS with numbers of margins and discussed about the self configuration of switching from UMTS to WiMAX architectures and vice versa to ensure the QoS required depending up on the network loads.

SIP signaling scheme with adaptive retransmission timer has minimum end to end delay with RLP for FER of 2% or greater. From [2],it can be concluded that zero buffer size is recommended but maximum playout interruption is to be avoided and PS network is out of imagination without buffer at nodes so de-jitter buffer of size 180 ms was evaluated at BLER of 2% with acceptable transmission delay with minimum playout interruption of 19.2 at 3.91 MoS greater than threshold of 3.5. RTCP can use DFS to control the data transmission rate reducing packet loss and end to end delay and TLS can be deployed for better throughput and improved jitter and delay. Form [10] higher FACH scheduling weight significantly reduces the End-to-End delay whereas for same scheduling weight, higher delay sensitive traffics are guided to dedicated channels (DCH) and less sensitive traffics by either shared channels or by dedicated channels. From [8] and [12], RSVP and DSCP can prioritize the traffic with higher delay sensitive so that delay sensitive class traffic is dequeued with higher forwarding behavior in all hops in the intermediate nodes in IP network.

III. NETWORK SIMULATION

Simulation of network was done with optimal CODEC [1]

TABLE 1. FACH Scheduling Weight

Parameter	Scheduling weight
Signaling	10.0
Conversational	7.0
Streaming	3.0
Interactive	2.0
Background	1.0

TABLE 2. UMTS to DiffServ Mapping

Application	UMTS Traffic	DiffServ DSCP
VoIP	Conversational	EF
Video	Streaming	AF1X
Web	Interactive	AF2X
FTP	Background	BE

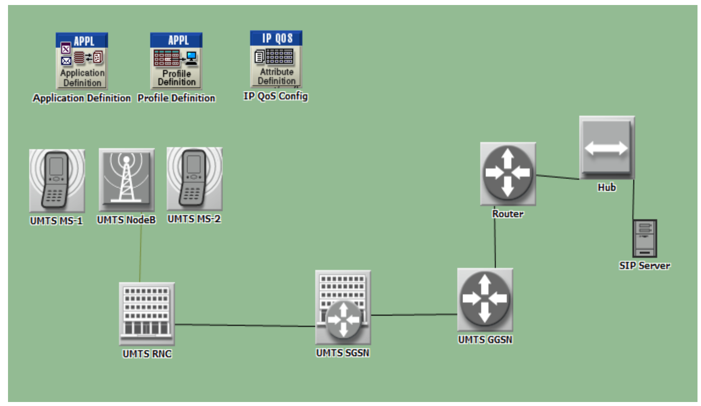


Fig.1 Simulation Model

along with appropriate frame size 20 ms and transport policies RSVP, FACH scheduling and DSCP mapping to UMTS QoS are studied in two scenarios using OPNET Modeler. In first scenario VoIP traffics over UMTS network with codec G.729A, RSVP enabled and FACH scheduling weight assigned as shown in table 1 for different traffics is simulated and the results are compared with that of second scenario where DSCP has been allocated to different types of traffics as shown in table 2 along with that of scenario first.

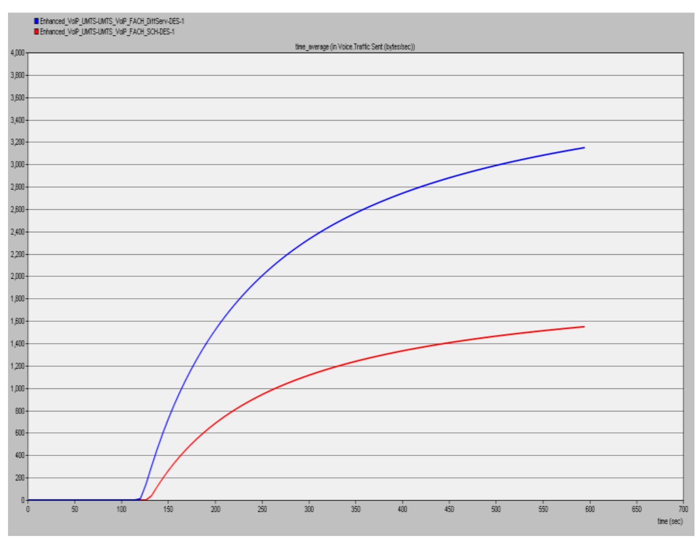


Fig. 2 Traffic Sent (bytes/sec)

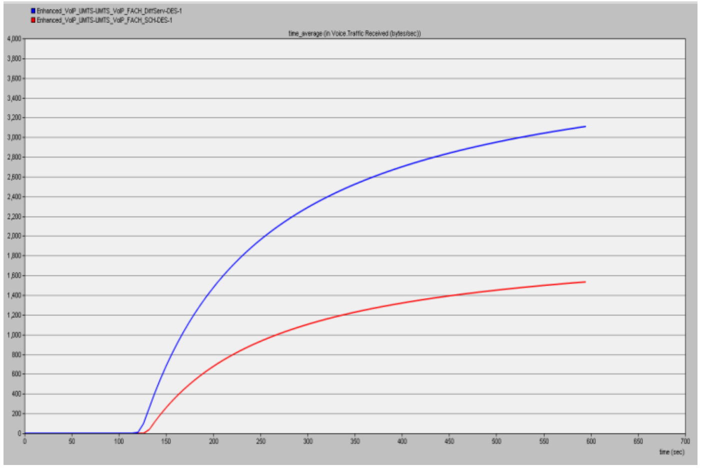


Fig.3 Traffic Received (bytes/sec)

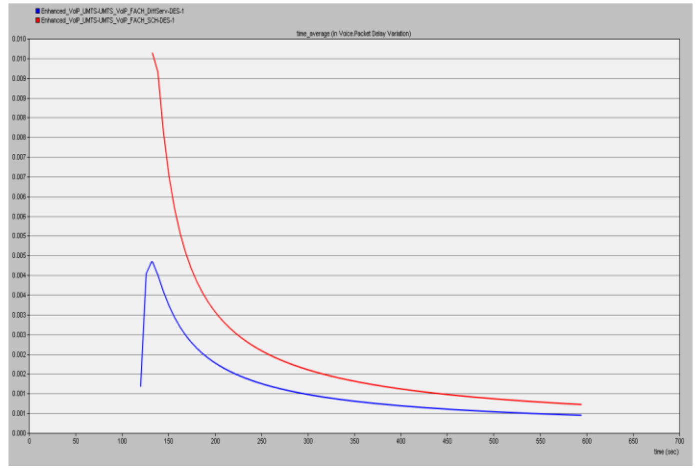


Fig. 5 Delay variation

IV. RESULT AND ANALAYSIS

Fig 2, 3, 4, 5, 6, 7 are some snapshots during simulation of networks in two different scenarios and their comparative studies. For overall observation, DSCP integrated network model has significant QoS enhancement for the VoIP service over the UMTS network.

From fig. 2, 3 and 4, clearly End-to-End delay is comparable even for almost two times higher traffic intensity in case of DSCP based QoS and similar End-to-End delay converge to similar range is due to the UMTS bearer service in UMTS backbone network. Observed End-to-End delay is in the range of 80 ms as in fig.4 which is as ITU-T recommendation to be less than 150 ms and from fig. 5, delay variation has significant enhancement in DSCP integrated scenario.

From fig. 2, 3 and 6, DSCP based UMTS bearer service has similar MOS value where both the network setup has

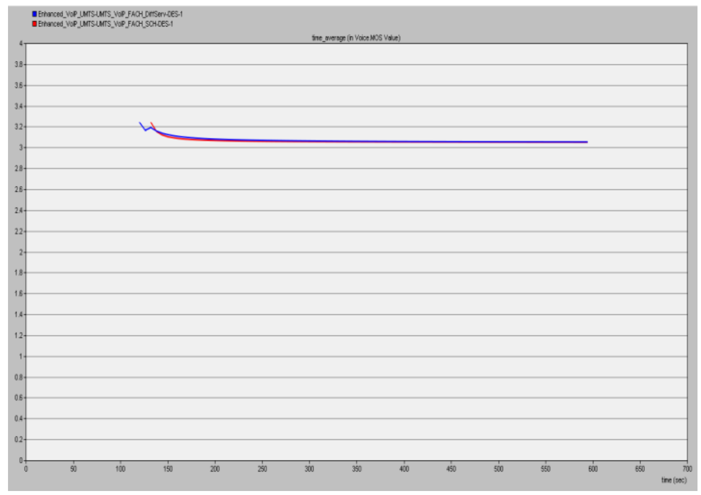


Fig.6 MOS Value

acceptable value of MOS greater than 3.0. Fig.7 indicates better jitter value for DSCP integrated network even for higher traffic intensity in UMTS network.

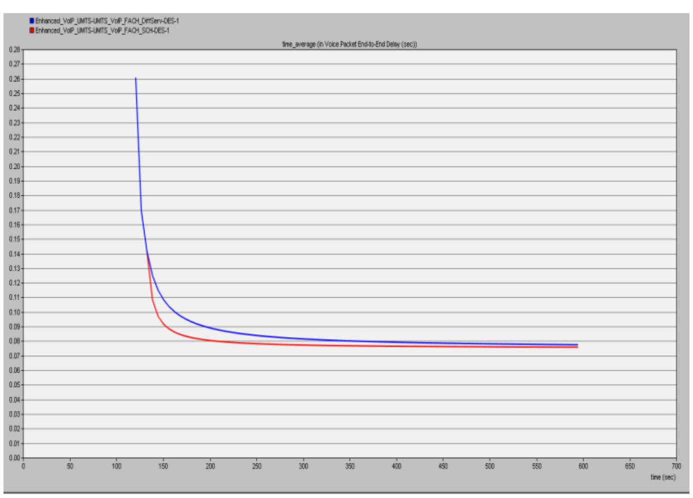


Fig. 4 End-to-End delay (sec)

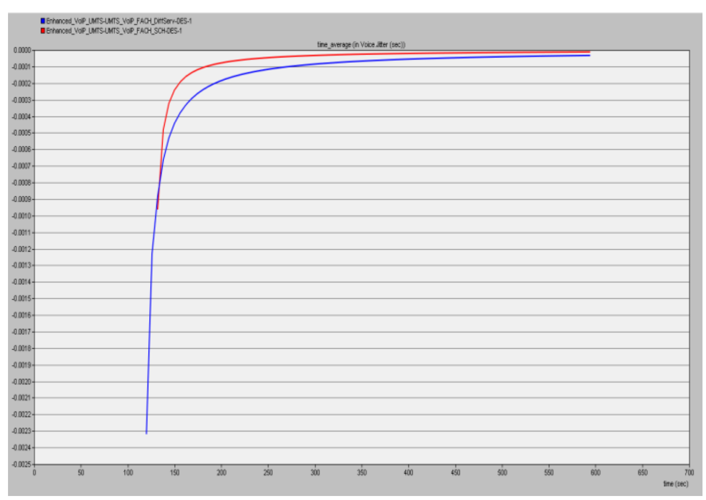


Fig. 7 Jitter (sec)

V. CONCLUSION

Better CODEC as in [1] G.729A with frame size of 1 ms will have good MOS and acceptable end to end delay. Delay can be minimized with FACH scheduling weight in UMTS Backbone network where conversational traffic is given higher scheduling weight than other streaming, interactive and background traffic over the UMTS network. Similar prioritized forwarding behavior in the external IP network can be negotiated by UMTS backbone network with DiffServ Code Point marked in packet header. Thus a QoS for VoIP can be modeled within UMTS Backbone network defining different UMTS bearer service to different class of traffics and in external IP network forwarding scheme with differentiated class of traffic with different DSCP allocated to different data traffic over IP network with significant improvement in End-to-End delay and better MOS though higher traffic load over the network.

REFERENCES

[1] J. Cao and M. Gregory, "Performance Evaluation of VoIP Services using Different CODECs over a UMTS Network," in Telecommunication and Networks and Applications Conference, ATNAC 2008, Australasia, pp. 67-71.

[2] A. Barbaresi and A. Mantovani, "Performance Evaluation of Quality of VoIP in UMTS-UTRAN R9," in communications ICC'07, IEEE International Conference, 2007, pp. 634-639.

[3] S. Jadhav, H. Zhang and Z.Huang, "Performance Evaluation of Quality of VoIP in WiMaX and UMTS," in proceedings of International conference on parallel and distributed computing, application and technologies, 2011, pp. 375-380.

[4] H. Fathi,S.S. Chakraborty and R. Prasad, "Optimization of SIP Session Setup Delay for VoIP in 3G Wireless Networks," IEEE Trans. on mobile computing, vol 5(9), pp. 1121-1132, September 2006.

[5] M. Sivabalakrishnan and D. Manjula, "Analysis of Decision Feedback Using RTCP for Multimedia Streaming over 3G," in proceeding of the International conference on computer and communication engineering, Kuala Lumpur, Malaysia, 2008, pp. 1023-1026.

[6] E.R. Vale, M. A. Grivet and J. C. Brandao, "Reducing the VoIP packet Overhead in UMTS Air Interface," in 2nd International Conference on Adaptive Science & Technology, 2009, pp. 92-98.

[7] R. Cuny and A. Lakaniemi, "VoIP in 3G Networks: An End-to-End Quality of Service Analysis," in 57th IEEE Semiannual vol (2) on Vehicular Technology Conference, 2003, pp. 930-934.

[8] R. Shreevastav, C. McGoldrik and M. Huggard, "Delivering Improved QoS and Cell Throughput in UMTS Based HSDPA Networks," in IEEE International Symposium on World of Wireless, Mobile and Multimedia Networks and Workshops, 2009, pp 1-9.

[9] 3Gpp "Technical Specification", Quality of Service (QoS) Concept and architecture, 3GPP TS 23.107 version 5.4.0 Release 5.

[10] M. B. Mallah and S. S. Islam, "Analysis the Impacts of Data Rates and Forward Access Channel Scheduling on QoS in 3G UMTS Network," in proceeding of the IEEE International Conference on Cyber Technology in Automation, Control and Intelligent System, Bankok Thailand, 2012, pp. 34-38.

[11] 3Gpp "Technical specification", Physical channel and mapping of transport channels on to physical channels (FDD), 3G TS 25.211 version 3.1.1 Release 1999.

[12] E. M. Saad, O. M. EL-Ghandour and M. K. Jehan, "Evaluation of QoS in UMTS Backbone Network using Differentiated Services," presented at 25th National Radio Science Conference (NRSC 2008), Egypt, 2008.

[13] S. V. Gaast, A. Hajjaoui and E. Meeuwissen, "Quality of Service for SIP Session in 3GPP-Based Networks," Bell Lab Technical Journal Vol.9(3), pp. 127-134, 2004.

Comparative Study of Integrated Transceiver for Real Time Monitoring in Rescue Operation

Md.Nasir Uddin*,MM Rashid,MG Mostafa,MJ Nayen
Department of Mechatronics Engineering
International Islamic University Malaysia (IIUM)
nasir.u@live.iium.edu.my, mahbub@iium.edu.my,
mostafauap@gmail.com, md.julkarnayen@gmail.com

[a]SZ Ahmed,[b]Sultan Mahmud,[a]NA Nithe,[c]JI Rony
Electrical & Electronics Engineering
[a]Atish Dipankar University of Science & Technology
zobair.ahmed@gmail.com, engnasirbd@yahoo.com
[b]Dhaka University of Eng. & Tech. sultan@duet.ac.bd
[c]Bhola Polytechnic Institute. engnasirbd@gmail.com

***Abstract*-An augmented reality system provides enhanced situational information to personnel located within an environment. A tracking system obtains viewpoint information corresponding to a real-time view of said environment. A processing system receives information from one or more sensors. Information includes sensor location information and status information about the environment and personnel therein. The processing system generates graphics using the sensor location information and the viewpoint information. Graphics include visual representations of said status information. A display displays the generated graphics on a display at a supervisor station that is outside of said environment such that graphics are superimposed on the real-time view.**

Keywords-Integrated Transceiver, Real Time Monitoring, Rescue Operation, Firefighter, Bluetooth, Wi-Fi, RF, Smart sensor platform, Motorola TETRA, SoC, GPS, Wireless.

I. INTRODUCTION

A Firefighter is a rescuer broadly trained in firefighting, largely to extinguish hazardous fires that threaten property, and to rescue people from dangerous situations such as collapsed or burning buildings and crashed vehicles. NIOSH (National Institute of Occupational Safety and Health) have conducted investigations on work-related firefighter deaths and statistics are as follows in figure 1.1. The National Fire Protection Association (NFPA) has conducted an analysis on the natures of duty associated with firefighter deaths. Contained within their report are the roots of deadly injuries to firefighters [1].

Figure 1.2 shows the causes of fire fighter fatalities. Based on the Figure 1.2 it clearly states that the frequent cause of injury is due to overexertion or stress which is 34%. This is a valid reason for monitoring firefighters physiologically. Real time assessment of the physiological status and the status of mounted device (oxygen level) pressure of the firefighter is very crucial to be monitored during a fire rescue operations. This real time assessment should be able to assess the baseline of physiological characteristics such as aerobic fitness, sleep history and heart rate. Figure 1.3 shows an overall view on communications method.

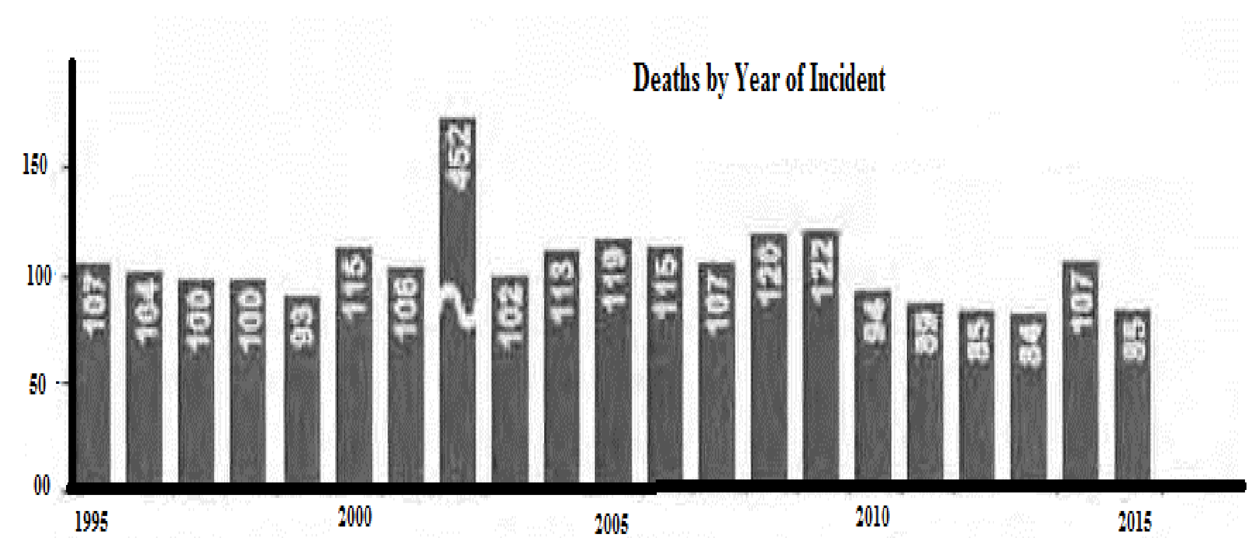


Fig.1: Firefighter Fatalities

II. ASSESMENT

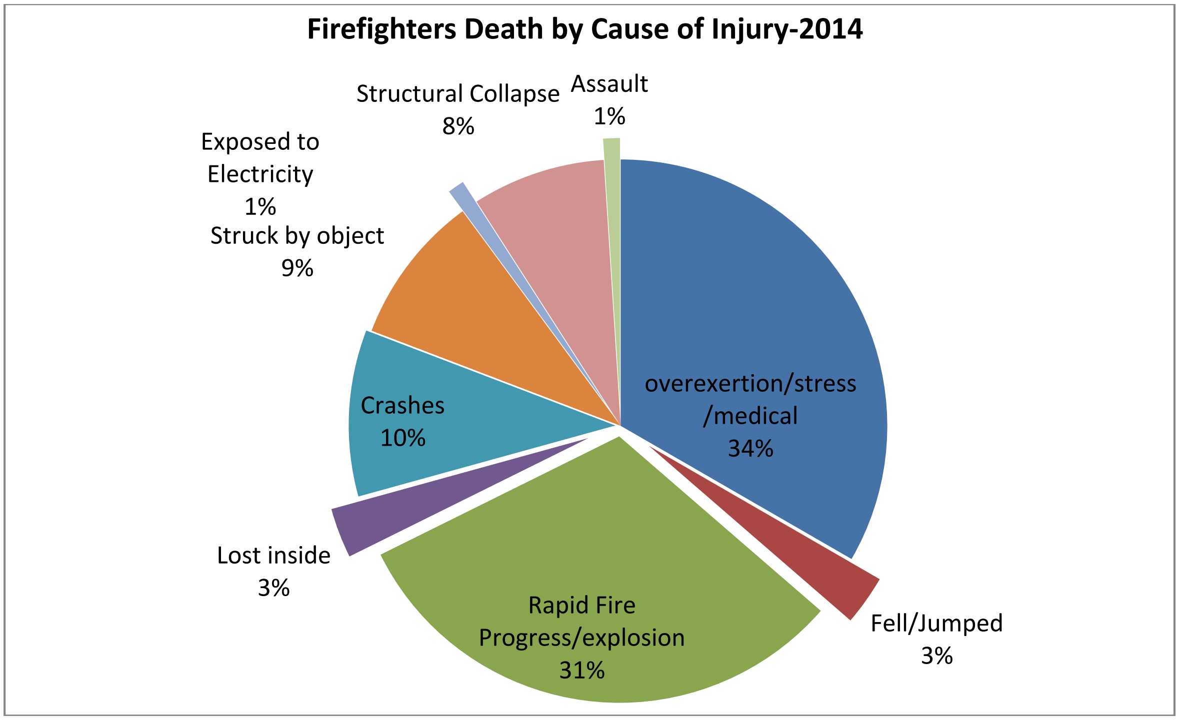


Fig.2: Causes of fire fighter fatalities.

It shows that firefighters which have been divided into two groups located in two different parts of the building reports back to the commander wirelessly via radio frequency. Apart from that, it should be able to provide mission support such as improve "who, where, when" situational awareness, guide acute and chronic work or rest cycles and reduce the likelihood environmentally related injures such as heat stroke which is quite common when it comes to fire rescue mission. Besides that, casualty evacuation to be facilitated and the quality of after action reviews to be improved. The following example provides an illustration of exemplary prior art systems.

It has long been desirable to provide enhanced situational awareness to first responders. For example, providing first responders with more information about their surrounding environment could improve rescue operations. Prior art devices have attempted to provide enhanced situational awareness to first responders by combining a virtual representation of an environment (e. g. a map or 3D representation of a building) With status information received from first responders and having a user interpret the relevance of the combination and communicate the relevance to first responders.

While the invention has been described in connection With What is presently considered to be the most practical and preferred embodiment, it is to be understood that the invention is not to be limited to the disclosed embodiment, but on the contrary, is intended to cover various modifications and equivalent arrangements included Within the spirit and scope of the appended claims [2]. Wherein the graphics include visual representations of the status information; displaying the generated graphics on a display at a supervisor station that is outside of said environment such that the generated graphics are superimposed on the real time view[3] that shows figure also [4]. The system can provide an enhanced representation of the environment displaying the generated graphics on a display such that the generated graphics are superimposed on the real-time view[5].This enhanced representation can be used to provide enhanced situational awareness for first responders on the real-time view[6] that shows the figure [7].

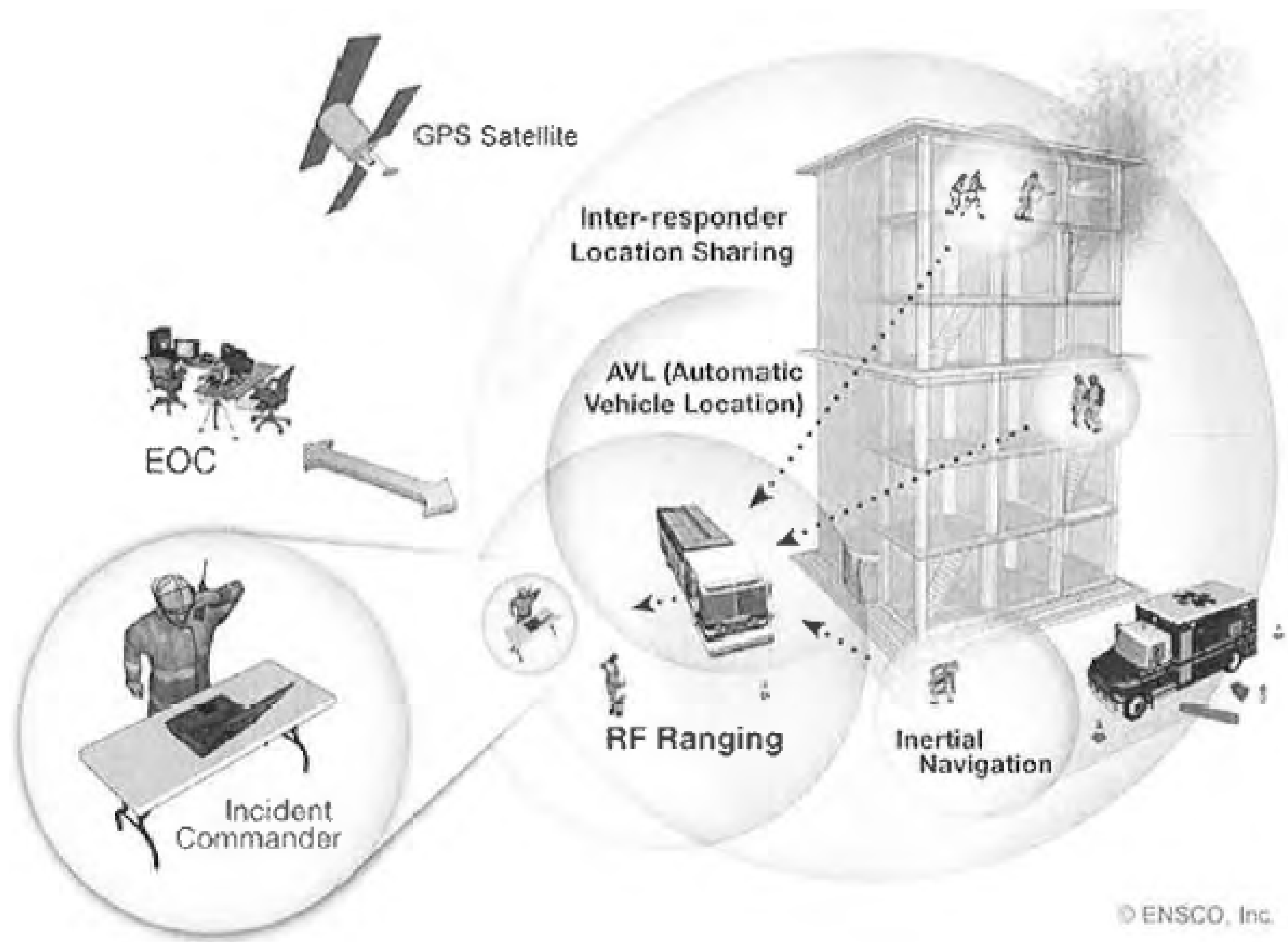


Fig.3: ENSCO Inc.'s Real Time Assessment

It has long been desirable to provide enhanced situational awareness to first responders. For example, providing first responders with more information about their surrounding environment could improve rescue operations. Prior art devices have attempted to provide enhanced situational awareness to first responders by combining a virtual representation of an environment (e. g. a map or 3D representation of a building) With status information received from first responders and having a user interpret the relevance of the combination and communicate the relevance to first responders.

While the invention has been described in connection With What is presently considered to be the most practical and preferred embodiment, it is to be understood that the invention is not to be limited to the disclosed embodiment, but on the contrary, is intended to cover various modifications and equivalent arrangements included Within the spirit and scope of the appended claims [2]. Wherein the graphics include visual representations of the status information; displaying the generated graphics on a display at a supervisor station that is outside of said environment such that the generated graphics are superimposed on the real time view[3] that shows figure also [4]. The system can provide an enhanced representation of the environment displaying the generated graphics on a display such that the generated graphics are superimposed on the real-time view[5].This enhanced representation can be used to provide enhanced situational awareness for first responders on the real-time view[6] that shows the figure [7].

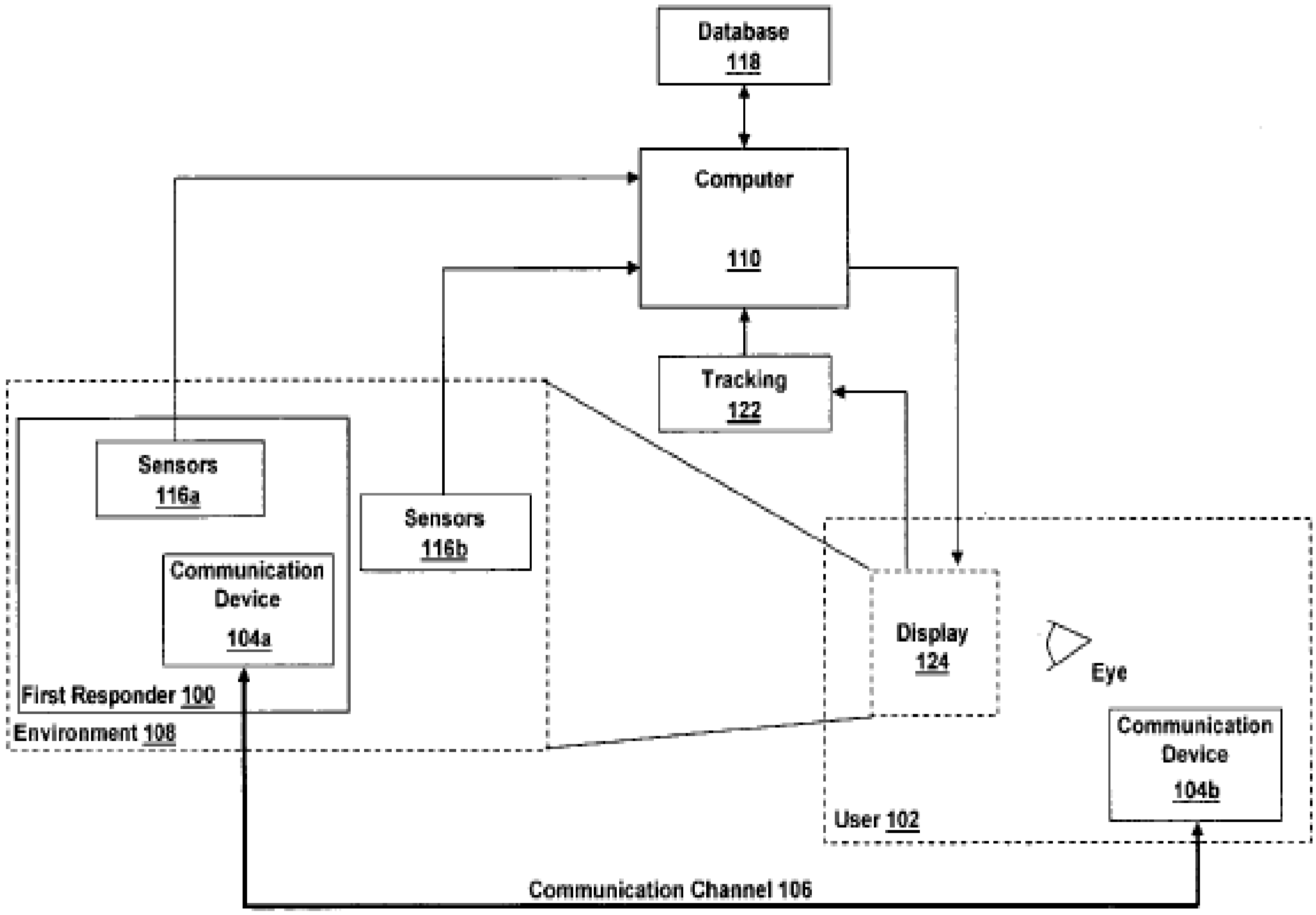


Fig.4: Basic Block of Real Time Assessment

The system can also show the locations and values of sensors placed within the environment superimposed on top of a real-time view of the environment. For example, when a temperature sensor is dropped by a firefighter, the sensor's own tracking system (or last location of the firefighter at the time he dropped the sensor) provides the location of the sensor. By showing data coming from the sensor on top of a real-time view of the environment, the captain can directly relate the sensor reading with a location in the environment [8].

Based on the Table 2.1 shows, Q Zi [9] has used Acquisition and communication module based on CC2430 as shown the usage of zigbee technology helps to transmit data from medical sensors to monitoring equipments via wireless transmission, which reduces the usage of cable links. There is a mini monitoring Network between the monitoring instruments and zigbee sensor nodes. The physiological information is detected using a controller which is installed on the sensor nodes then the information will be transmitted via wireless transmission to the selected equipment[10]. Communication transmission system can be monitored for 24 hours due to its low power consumption . Apart from that, wireless sensor network or WSN is a wireless network that comprise of independent devices such as low power consuming processor, Flash memory, ADC, RF Transceiver [11].

Feedback can be received and monitored remotely via wearable monitoring system. A wearable system is used to detect the information and it consist if data collection hardware, remote centre and data analysis [12]. Based on another article which uses SoC or System on Chip platform together with Bluetooth wireless network. Bluetooth module is set as the transmitter where as the SoC is set as receiver platform [13]. Besides that according to the article of Ramamurthy, Pravu and Rajit a smart sensor platform which is based on a patent pending technologies has a plug and play proficiencies [14]. This supports hardware interfaces and communication desires for many sensors [15].

III. OVERVIEW

Table 2.1 shows the related work in this area of research.

No	Authors	Title	Year	Platform	Communication	Merits	Demerits
1	Qi Zhao, Qi Wu, Jianbo Wu, Xiaoming Wu	Design of Physiological parameter Acquisition & Communication Module Based on CC2430	2008	Zigbee, CC2430	Zigbee	Low Power Usage, Bulky	No means of radio Call
2	Jin Soo CHOI, MengChu ZHOU	Recent Advances in wireless sensor Networks for Health Monitoring	2010	Wireless Sensor Network	Zigbee, Bluetooth, WiFi	Low Power Usage, Bulky	No means of radio Call
3	Jzau-Sheng Lin, Shi-Yuang Huang, Keo Wen Pan, Shao-Han Liu	A Physiological signal monitoring system based on an Soc Platform & wireless network technologies in Homecare Technology	2009	SoC (System on Chip)	Bluetooth	Low Power Usage, Bulky	No means of radio Call
4	Motorola Solutions	Motorola Tetra Terminals	2015	TETRA Portable Terminal	Bluetooth, WiFi, RF	Portable Radio Call Available, Programmable	Not Available
5	P.S Pandian	Wireless sensor network for wearable Physiological monitoring	2008	wireless sensor Networks	Bluetooth, WiFi	Bulky	Radio Call not Available
6	Shyamal Patel, Hyung Park, Paolo Bonato, Leighton Chan, Mary Rodgers	A Review of wearable sensors & System with Application in Rehabilitation	2012	SoC, wireless sensor Networks	Bluetooth, WLAN, Zigbee	Low Power Consumption, Wearable	Radio Call not Available
7	Harish Ramamurthy, B.S Prabhu, Rajit Gadh	Wireless Industrial Monitoring & Control using a smart sensor platform	2007	smart sensor platform	Bluetooth, WiFi	Plug & Play	Radio Call not Available
8	Rae Systems Inc.	Using Wirelessly connected Monitoring Equipments	2011	Wireless Connected System	Bluetooth, WLAN	Fast & Flexible Deployment	Radio Call not Available
9	Keith Ammons	Pros and cons of Tetra Vs P25 & the benefits of a multi technology platform for Tetra, P25 Phase I/Phase Ii & Mobile Wimax	2014	Power Trunk Tetra	Bluetooth, WiFi, RF	Enhanced for high Population density Area	Not currently offered in VHF band
10	Motorola Ltd.	Motorola Tetra Solutions	2009	Motorola Tetra	Bluetooth, WiFi, RF	Multi Slot Packet Data, Radio Calls at Long range	Not Available

IV. CONCLUSION

Another platform that was introduced as Motorola Tetra portable terminal which provides [16]anupright functionally combined with GPS and nonstop encryption [17]. TETRA is an open standard for digital radio mobile communication and its also known as Terrestrial Trunked Radio [18]. TETRA terminal has quick access to voice & data service, besides that it has programmable interface which can be programmed according to the user. It also has radio call in the frequency band of 380-400 MHZ, 806-870 MHZ spectrum and RF channel Bandwidth of 25 KHZ[19]. It provides convenience and hands free mobility. It can be clearly seen that Motorola Tetra has a better perform platform compared to others. A high durability, portable device integrated with radio call has better function compared to other platforms in a field work [20]. The current system that is being used in Malaysia by [21]is still conservative where real-time assessment is not done. The risk of each firefighter who is exposing their lives in a fire rescue operation is still presence. With the basic walkie-talkie communication between firefighters is not sufficient to assure their safety during the high risk fire rescue operation[22].

REFERRENCES

[1] N. F. P. Association, *NFPA 1500, standard on fire department occupational safety and health program*, 2012.

[2] J. D. Courtney, "Monitoring system having wireless remote viewing and control," ed: Google Patents, 2002.

[3] J. D. Snell, "Method and apparatus for monitoring physiological data from an implantable medical device," ed: Google Patents, 1998.

[4] G. R. Lawson, C. D. Harper, G. M. Brehm, and G. S. Kalyanpur, "System and method for monitoring service quality in a communications network," ed: Google Patents, 2002.

[5] F. S. Groenteman, "Office machine monitoring device," ed: Google Patents, 1998.

[6] E. Camhi, "System for monitoring a subject," ed: Google Patents, 1998.

[7] R. Bary and J. Rialan, "Seismic acquisition and transmission system with functions decentralization," ed: Google Patents, 1998.

[8] W.-S. Jang, W. M. Healy, and M. J. Skibniewski, "Wireless sensor networks as part of a web-based building environmental monitoring system," *Automation in Construction,* vol. 17, pp. 729-736, 2008.

[9] Q. Zhao, K. Wu, J. Wu, and X. Wu, "Design of Physiological Parameter Acquisition and Communication Module Based on CC2430," in *7th Asian-Pacific Conference on Medical and Biological Engineering*, 2008, pp. 348-351.

[10] J. S. Choi and M. Zhou, "Recent advances in wireless sensor networks for health monitoring," *Int. J. Intell. Control Syst,* vol. 14, pp. 49-58, 2010.

[11] J.-S. Lin, S.-Y. Huang, K.-W. Pan, and S.-H. Liu, "A physiological signal monitoring system based on an SoC platform and wireless network technologies in homecare technology," *Journal of Medical and Biological Engineering,* vol. 29, pp. 47-51, 2009.

[12] A. K. Salkintzis, "Evolving public safety communication systems by integrating WLAN and TETRA networks," *Communications Magazine, IEEE,* vol. 44, pp. 38-46, 2006.

[13] P. S. Pandian, K. P. Safeer, P. Gupta, D. Shakunthala, B. Sundersheshu, and V. C. Padaki, "Wireless sensor network for wearable physiological monitoring," *Journal of networks,* vol. 3, pp. 21-29, 2008.

[14] H. Ramamurthy, B. Prabhu, R. Gadh, and A. M. Madni, "Wireless industrial monitoring and control using a smart sensor platform," *Sensors Journal, IEEE,* vol. 7, pp. 611-618, 2007.

[15] A. Mainwaring, D. Culler, J. Polastre, R. Szewczyk, and J. Anderson, "Wireless sensor networks for habitat monitoring," in *Proceedings of the 1st ACM international workshop on Wireless sensor networks and applications*, 2002, pp. 88-97.

[16] L. C. Lowe, S. G. Senaratne, and K. W. Colston, "BONE MARROW STROMAL CELL-DERIVED OSTEOPROTEGER IN (OPG) PROTECTS BREAST CANCER CEL LS FROM TRAIL-IN DUCED APOPTOSIS," *British journal of cancer,* vol. 88, pp. S25-S54, 2003.

[17] X. UltraCell's, "Fuel cell backup for TETRA base stations in Denmark," *Fuel Cells Bulletin,* 2009.

[18] F. D. Alotaibi and A. A. Ali, "TETRA outdoor large-scale received signal prediction model in riyadh city-saudi arabia," in *Wireless and Microwave Technology Conference, 2006. WAMICON'06. IEEE Annual*, 2006, pp. 1-5.

[19] O. Grossman, M. Ben-Ayun, and S. Yaniv, "Multi-frequency band antenna and methods of tuning and manufacture," ed: Google Patents, 2004.

[20] M. Rantama and K. Junttila, "Pelastustoimen langattoman tiedonsiirron tarpeet ja toteutusmahdollisuudet tulevaisuudessa," *Pelastusopiston julkaisu. B-sarja: Tutkimusraportit,* vol. 2, p. 105, 2011.

[21] M. S. Aini, A. Fakhru'l-Razi, and M. Daud, "Evolution of emergency management in Malaysia," *Journal of Contingencies and Crisis Management,* vol. 9, pp. 46-53, 2001.

[22] N. H. Salleh and A. G. Ahmad, "Fire safety management in heritage buildings: the current scenario in Malaysia," 2009.

Development of An Absorption Silencer for Generator's Noise Reducing

Md.Nasir Uddin*,MM Rashid,MG Mostafa,MJ Nayen
Department of Mechatronics Engineering
International Islamic University Malaysia (IIUM)
nasir.u@live.iium.edu.my, mahbub@iium.edu.my,
mostafauap@gmail.com, md.julkarnayen@gmail.com

[a]SZ Ahmed,[b]Sultan Mahmud,[a]NA Nithe,[c]JI Rony
Electrical & Electronics Engineering
[a]Atish Dipankar University of Science & Technology
zobair.ahmed@gmail.com, engnasirbd@yahoo.com
[b] Dhaka University of Eng. & Tech. sultan@duet.ac.bd
[c] Bhola Polytechnic Institute. engnasirbd@gmail.com

***Abstract*-Noise pollution is considered to be one of the major environment pollutants which affect human beings both physically and psychologically, as such, a noise-free environment is in great demand worldwide. Diesel engine generators are highly appreciated as power sources of electric equipment in factories, houses and business centers. Loud sounds from diesel generators are a major cause of noise pollution. This paper analyzes the noise source of diesel generators and mitigates this pollution by a modified absorbance silencer or muffler. For automotive engines, the principle source of noise is its intake, radiator, combustion, etc. In our society, all of the industries, the residential sector and business plants use generators. In this research, an absorbance silencer is modified for reduced noise of the generator set. It is constructed from a combination of baffle or perforated duct with sheet metal. This paper aims to study the sound characteristics of generator sets and also aims to reduce the sound by means of a well-modified muffle silencer. This paper focuses on design and tests silencers, particularly absorption silencers for engine exhausts.**

Keywords: Diesel Engine; Generator; Absorption Silencer; Noise.

I. INTRODUCTION

Sound pollution means unwanted sounds or noise. It is perceived by most people as annoying. Noise pollution harms most people's lives. Additionally, it is a great cause of environmental pollution. It greatly hampers humans not only physically, but mentally also. For these reasons, noise reduction is in great demand in this society, and noise prevention is a rising concern in all markets. In our society, all of the industries, the residential sector and business plants use generators. Diesel engine generators cause loud sounds.

A silencer is essential and an important part for sound attenuation of engine exhausts. There are many theories and designs of acoustic silencers of generator sets, developed in detail by Stewart theory and design of Acoustic and silencer of Generator set developed in detail by Stewart [1, 2] and he apply it to create many types of silencer and also success that explained in[2]. In 2013 Dr. Chazot, Nenning and Perry performed the method of unity finite element of 2D noise field with sound absorbing materials [3]. Now a days Ontop is a large company who designing, producing and manufacturing prefab modular flue and also distributing. It disposes of a modern product that certified ISO 9001 and also environment friendly as metaloterm lightweight silencer for flue system. In 2012, Mr. Ghosh, Bose and Chakraborty in india modified muffler and get a good performance of a diesel engine by used it [4]. The review of Generator set and silencer should be not complete without it mixed the effects of different absorption elements [5] .The diesel engine is the main noise sources of sound power also the generator exhaust and radiator fan [6], are measured by the method of sound intensity. At first May and Olson expressed an electronic noise absorber by pressure release on back face of resistive sheet [5]. Its introduce the notion of active absorption. Guicking and Lorenz in the year of 1980 fulfilled this theory and done experiment [7]. Various research have sought to complement multiple hybrid absorbance technique, leading to patent application [8]. In 1997 Mr. nail and Furstoss improved a layer of optical wool backed by air cavity closed through an active surface [9] by an active treatment. In the same year Beyene and Burdisso found active boundary condition [10]. They achieved it by impedance adaption means in layer of porous rear face.

But after the century in 2004 cobo et al. explained structure of thinner hybrid active and passive absorbers feasibility. He used micro perforated panels more than the porous materials[11]. The design mufflers and procedures are also in the literature (Munjal, 1987)[12]. Long time ago Stewart used electro acoustic analogies in deriving acoustic filters theory & design.[1]. After that Davis approach systematic studies result of muffler.[13]. Igarashi and M. Toyama calculated transmission characteristics by using electric circuit. [14, 15]. The last year in 2014 Babu, A.R Rao simulated a new muffler for reduce sound level of SI engine.[16]

In this paper, an absorbance silencer is modified for reduced noise of a generator set. It is constructed using a combination of baffle or perforated duct with sheet metal. The maximum generator has a simple silencer for reduction of the exhaust noise only.

Noise means unwanted sounds that are abnormal or loud- it is a relative term. Singing or hearing a song or musical instruments may be noise for some. Automotive engines create a large portion of the noise in our society. I.C engines are also a great source of sound pollution, as they are a powerful source of noise. The noise sources of both gasoline and diesel engines are the same, but their noise characteristics are different. Noise is highly subjective, and that which is irritating to one can be acceptable for another. To overcome this, noise is measured by a decibel (dB) meter in unit of dBs, withdB(A) representing the human ear's sensitivity of 0 to 180 dB, where 0dB means no sound at all, and 180dB is a loud sound. An alternative explanation for 180 dB is the level of sound an atomic bomb would make upon explosion. The dB scale is a logarithmic meter. If dBs risein increments of 10, then the sound level rises 10 decade or 10 fold. If we know the level of noise source and maximum allowed level, then it is easy to calculate the required sound reduction for the silencer. The USA standard ASTM E413 describes frequencies of machinery as being in the range of 125 to 4000 Hz [17]. Similarly, The international standard ISO 717 refers to frequencies 100 to 3150 Hz[18]. The SI unit of sound reduction is dB and frequency is Hz[19].

II TECHNIQUES

The noise reduction techniques are dependent on the generator room, exhaust and its type of structure, borne, noise & vibration. Some techniques are shown in the following-

***A. Generator Rooms*:**

a. Room Enclosure:

i. Roof

ii. Walls

iii. Doors

iv. Internal Lining

b. Intake Air andDischarge Air:

i. Duct Silencers

ii. Acoustic Louvers

iii. Exterior Screens

***B. Exhaust Noise*:**

a. Resistive Mufflers/ Absorbance Silencer

b. Active Noise Control

***C. Structure Borne Noise & Vibration*:**

a. Spring isolators on generators larger than 175kW.

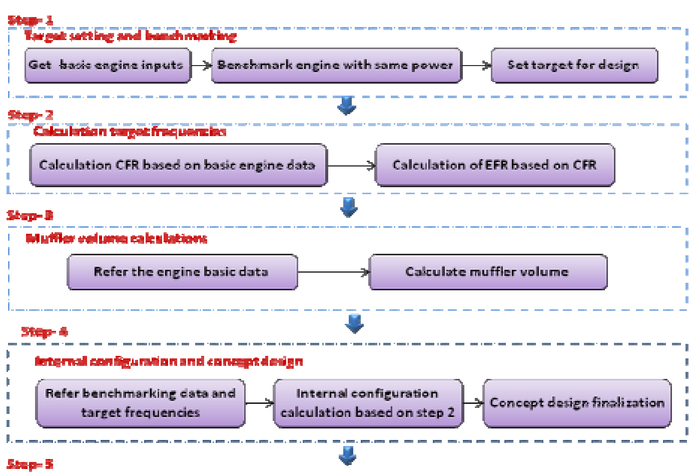
b. If a floor joint is present, the weight of concretebeneath the generator should be not less than twicethe generator weight.

c. Flexible pipe connectors, duct connectors,electrical connection at the generator.

Active noise cancellation silencers used to be available as amanufactured product, but are not currently available. Theywere effective in reducing the low frequency tones associatedwith the cylinder firing. In this research paper, we design and modified resistive mufflers / absorbance silencers for reduction of exhaust noise.

III METHODOLOGY

The methodology involves silencer design and development, and consists of some steps. After this, a modified silencer for use with a generator for a practical experiment is produced.The properly designed muffler for any particular application should satisfy the often – conflicting demands of at least five criteria simultaneously. ***A. Criterion and Flowchart of Methodology*:**The acoustic criterion, which specifies the minimum noise reduction, is required from the muffler as a function of frequency. The operating conditions must be known because large steady-flow velocities or large alternating velocities (high sound pressure levels) may alter its acoustic performance. The aerodynamic criterion specifies the maximum acceptable average pressure drop through the muffler at a given temperature and mass flow. The geometrical criterion specifies the maximum allowable volume and restrictions on shape. The mechanical criterion may specify materials thatare durable and require little maintenance. The economical criterion is vital in the marketplace [33]



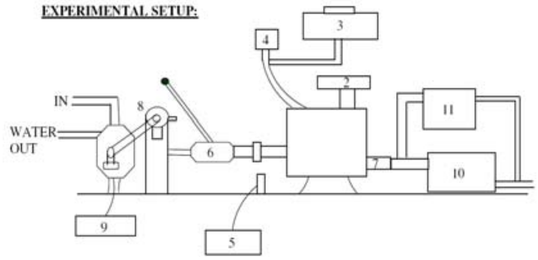


Figure & Step 1 to 6. The steps showing the process of design of the silencer andthe experimental setupof the generator set with a diesel engine.

1. Diesel Engine
2. Filter
3. Tank
4. Burret
5. RPM Indicator
6. Clatch or Shaft
7. Exhaust Outlet
8. Alternator
9. Radiator
10. Silencer
11. Sound Meter

The generator block diagram is replaced by the experimental setup block diagram. The various types of generator sets include 150KW, 350KW, 500KW, 1MW and 2MW diesel engines for use during the experiment and data collection. The experimental silencer was designed for a 500KW diesel engine generator set, and the basic specifications of the generator set are given in Table 1.

Table 1. Specifications of the generator set.

SN	Item	Specification
1	Rating	635KVA
2	Power	508KW
3	Current	850A
4	Rated revolution	1800RPM
5	Pressure	460KPA
6	cylinder	6
7	Cycle/stroke	4
8	Engine Load	75% and Full Load also

B. Experimental Evaluation of Unsilenced exhaust Noise

The noise of an engine exhaust varies significantly with its loading. At the full load, the sound level is about 10 dB more than the no-load condition. The silenced exhaust noise level is high at low frequencies. Figure 7 shows a 2MW engine with unsilenced exhaust noise level load for 16cyl at 1800RPM.

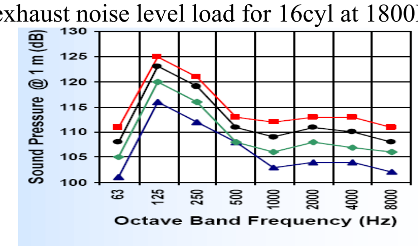


Figure 7. Sound Pressure at 1m distance for 2MW Engine by load.

The graph shows that the exhaust system starts at 110dBA and varies by 10 dBA, reaching a maximum of 120dBA. It is measured 1m from the outlet exhaust. The exhaust sound is affected by turbochargers of engines and after coolers by cooling fans. Hence, collecting noise data from engines is the optimal method chosen for this experiment. The unsilenced engine's exhaust noise level is high at low frequency. Figure 8 shows data comparisons for the various engines including 150KW, 350KW, 500KW and 2 MW diesel engines.

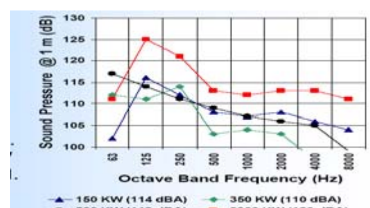


Figure 8. Sound Pressure of various Engine at 1m distance.

The spectrum of exhaust noise always contains loud sound associated with the cylinder firing rate (CFR). Each cylinder fires once every drive shaft revolution in a 4-cycle engine, and the CFR is calculated with different formulas for 4 cycle engines (Equation (1)), and 2 cycle engines (Equation (2)).

$$CFR = \frac{RPM}{120} \quad (1)$$

$$CFR = \frac{RPM}{60} \quad (2)$$

The engine firing rate is defined as

$$EFR = N * CFR \quad (3)$$

Where, N= number of cylinders.

Figure 9 shows the exhaust noise of a 500 KW diesel engine with 6 cylinders, running at 1800RPM without using any silencers. The narrow band spectrum data was collected at a 3m distance from the outlet of the exhaust without use of a silencer, with the enginerunning at full load. The engine firing rate (EFR) is 90Hz and the CFR is 15 Hz.[21]

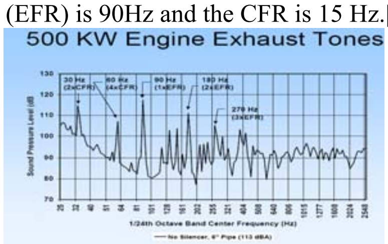


Figure 9. Sound Pressure Level of CFR & EFR without silencer

IV DESIGN & PRINCIPLES

The first step in any design and development activity is to set a target by doing a benchmarking exercise of models, which was carried out in this experiment.

A. Benchmarking:

After the benchmarking exercise, one needs to calculate the target frequencies to give more concentration of higher transmission loss. The primary step in silencer design is benchmarking based on engine input data:

$Bore, D = 80mm$

$Stroke, L = 98mm$

$Number\ of\ Cylinder, n = 6$

$Engine\ power, P = 65hp$

$Max. RPM = 1800RPM$

$Allowable\ Back\ Pressure = 10\ in\ H2O$

$Transmission\ Loss\ Noise\ target = 30dB$

B. Calculation of CFR & EFR:

The exhaust noise always contains loud sounds associated with the CFR. Each cylinder fires once every drive shaft revolution in a 4-cycle engine, as can be seen in Equation (1) and (2).

$CFR = \frac{1800}{120}$=*15Hz.*

Engine Firing rate (using Equation (3)): *6*15=90Hz*

C. Volume of the muffler (Vm):

The volume of the muffler isdefined as Vm, with units in litres. The calculation of the volume can be done using Equation (4):

$$Vm = Vf * \frac{\pi}{4}(d^2 * l) * (\frac{n}{2}) \quad (4)$$

Swept volume per cylinder is calculated as follows:

$$Vs = \frac{\pi}{4}(d^2 * l) = \frac{3.14*80^2*98}{4} = 0.5\ Lit. \quad (5)$$

Total *n*Vs=6*0.5=3 Lit.*

*Volume, Vm=(n)**$\frac{Vs}{2}$*=1.5 Lit.*

The silencer volume is considered to be at least 12 to 25 times larger, with a factor of 16

$$Silencer\ Volume = 16 * 1.5 = 24\ Lit.$$

D. Insertion Loss:

Figure 10 shows insertion loss for various mufflers, showing the absorptive muffler performance being optimal in the frequency region of 1000Hz to 2000Hz.

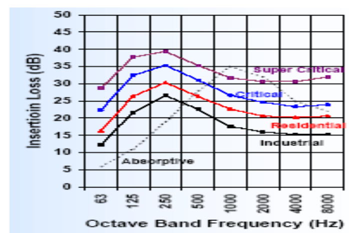


Figure 10. Insertion Loss of various Muffler

The 500 KW generator engines have an unsilenced exhaust noise level (UNL) of 116 dBA at a 1 m distance. Asafety factor (SF) of 5 dBa is allowed for noise transmission paths. The Exhaust noise criteria (ENC) = Required Noise Criteria (RNC) -5 dBA. This means that if our expected noise level is 60dBA, then we have to design a muffler for 55dBa. The UNL equation from the exact location is shown in Equation (6):

$$Lp(xr) = Lp(x0) - 20\ log\ (xr/x0) \quad (6)$$

For example, $Lp(25\ m) = Lp(1\ m) - 20\ log\ (25/1)$

$Lp(25\ m) = 116 - 28 = 88\ dBA$

$The\ required\ insertion\ loss, IL = UNL - ENC + SF.$

$IL = 88\ dBA - 55\ dBA + 5 = 38\ dBA.$

A silencer element'stransfer matrix method (TMM) is a function of state variables [28] , geometry of elements, velocity of mean flow, duct liner properties [29]. The transfer matrix method also influenced by temperature, nonlinear effects, high order mode etc.[30]. The Transmission Loss is shown in Equation (7) below.[24, 31, 32]

$$TL = 20\log\left|\frac{T_{11}+T_{12}/Y_1+Y_{11}T_{21}+T_{22}}{2}\right| \quad (7)$$

E. Internal Configuration and Design of the Proposed Silencer:

The silencer contains glass wool shielded from the exhaust stream by perforated metal. Glass silk, fiber optic or steel wool is commonly used. When the absorbance silencer works effectively, the materials suffer from deterioration in service. The combustion products take the form of absorbing materials. Materials melt due to heat generation until they have low thermal conductivity. The absorbance silencer is designed with low pass filter forms in order for it to be able to deal with the low frequency. Effective measures were used to reduce the sound. The noise power has to be applied in the numerical analysis.

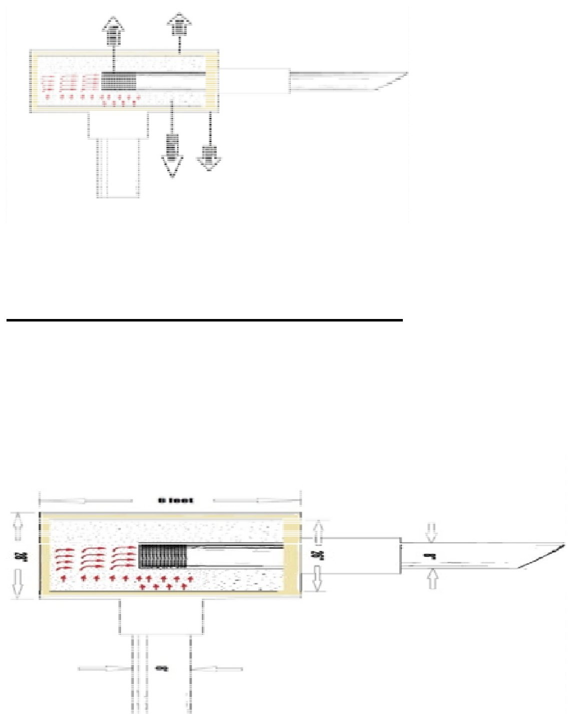


Figure 11(A,B). Design of Modified Absorption Silencer.

The operation and principle of the new absorption silencer is shown in Figure 11. Exhaust gas enters from the inlet pipe and is directed in multiple directions in the indoor chamber. The indoor space has a U-shape configuration with large spaces. Therefore the gases flowing into the space from the inlet to the outlet are distributed by the inner pipe hole. The inner pipe also has absorption materials like glass fiber, steel wool and sheet hole. The exhaust gases are absorbed automatically by these materials as they move around the inner space. The flow of these gases interfere with the leading gas flow, causing it to have a lower speed [23]. Figure 12 shows the inlet pipe and tail pipes (outlet pipes) with a diameter of 8 inches. The main perforated chamber is 6 feet long with a 28inch diameter. The absorption materials on the coating layer are only 2inches wide.The exhaust outlet pipe has resonance that increases its noise.To remove this,a short tail was used with a length of a quarter wavelength($\lambda/4$). Equation (8) describes the size of the tail pipe that described by jerry j lilly in AGL acoustic [17].

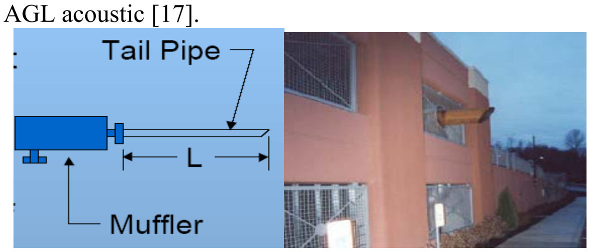


Figure 12. Tail Pipe.

Here n= positive integer number. when $L = n\lambda/2$, then occurs Resonance and for this reason this size is avoided. Resonance frequency of Tail pipe,

$$fn = nc/(2L) \tag{8}$$

c = sound speed. For a four stroke engine the EFR frequency is 90Hz and its wavelength is 20ft. The best tail pipe is exactly 5 ft. for cancel the EFR frequency of 90 Hz tone at the exhaust of outlet [21]. Here give the calculation for 6 cylinders @ 1800 RPM (950^{0}F)

$$CFR = \frac{1800}{120} = 15\,Hz$$

$$EFR = 6CFR = 90\,Hz$$

$$c = 49.03 * \sqrt{(460 + 950)} = 1841\ ft/\sec$$

$$\lambda_{CFR} = \frac{1841}{15} = 122\ ft.$$

$$\lambda_{EFR} = \frac{1841}{90} = 20\ ft.$$

$$L = \frac{20}{4} = 5\ ft.$$

Where L= tail pipe length. The tail pipe is a metal sheet that lies downstream of the exhaust silencer and has an acoustic resonance that can increase or amplify the final exhaust noise if matched. This resonance can be removed by making the tail half of the wavelength at the tone or sound frequency. However, it is advisable to avoid the tone by creating an accurate size at a quarter of the wavelength. The pipe hole's or perforated holes' number and diameter with measurementsare given in Figure 13 [24].

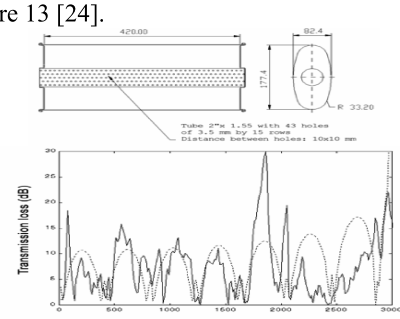


Figure 13: Transmission Loss of Concentric Perforated Tube.

The pipe hole of expansion chamber of the inner space helps to reduce the sound. The inlet pipe and outlet pipe can be extended to get more attenuation.[25, 26]. The absorption materials also reduce higher frequencies, especially that of mineral wool or glass wool [27].

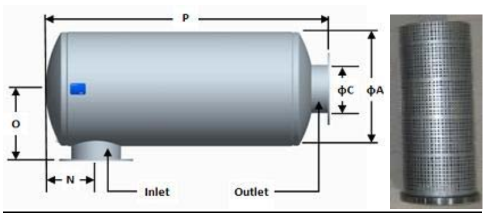


Figure 14. Side View of Absorption Silencer and Perforated hole

The diameter of inlet and Outlet exhaust pipe is-

$$Vm = \frac{\pi}{4}(d^2 * l)$$

D2=0.04

D=0.2m=200mm

And the perforated hole diameter is, $d_1 = \frac{1.29}{\sqrt{N}}$.[22]

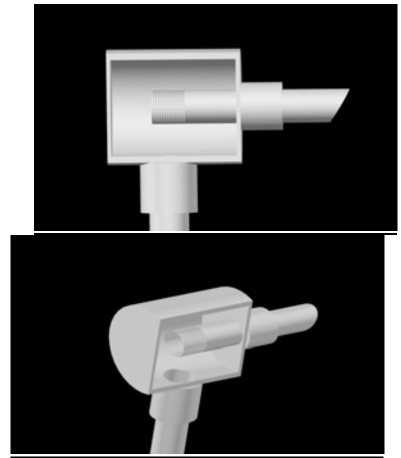


Figure 15. 3D view of Modified Absorption Silencer.

V RESULTS

The silencer design is successfulas it reduced the overall noise to the lowest level that can be reachedwithin acceptable limits. It is of good quality and does not have any effect on engine performance. The noise or sound attenuation characteristics of the new absorption silencer was measured andalso compared with the old silencer and is presented in Table 2. Shao (2011) measured and tested a new muffler and compared it with traditional muffler. The new muffler was designed with a combination of absorbance materials, a perforated pipe, an expansion chamber, a buffle and interpole ducting [23]. Figure 16 shows the test result.

Table 2. Sound attenuation characteristics.

SN	DISTANCE FROM SILENCER	PREVIOUS RECORD dBA	AFTER RECORD dBA	GENERATOR LOAD	PREVIOUS TEMP *C	AFTER TEMP *C	PRESURE KPA	RPM
01	1 Meter	120 dBA	85 dBA	75 %	82* C	82* C	460	1800
02	2 Meter	109 dBA	80 dBA	75 %	82* C	82* C	460	1800
03	3 Meter	106 dBA	70 dBA	75 %	82* C	82* C	460	1800

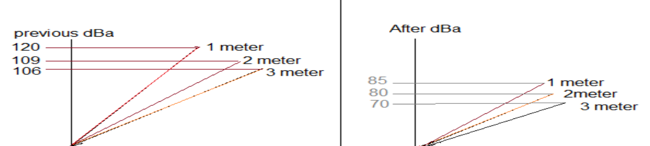


Figure 16. Sound Test Result.

Figure 16 shows that the sound pressure level decreases by approximately 30dB with a modified absorption silencer as compared to a traditional silencer. It also gives a better performance at various distances from the outlet exhaust as compared to other silencers. At 1500 RPM, the modified silencer gives the best result without any change of engine parameters -for example the temperature, pressure and KPA is the same as other traditional silencers. Figures 17 - 19 show the level of sound pressure of an exhaust in three types of silencers. Table 2 and Figures 17 - 19 illustrate that the modified absorption silencer has better noise reduction properties than other, traditional silencers and mufflers. Figure 20 shows a narrow band spectrum data, collected from a 3m distance from the outlet of an exhaust, used with a proposed silencer. Note the dip in the curve in the vicinity of 80 Hz and 240 Hz. The fact that there is no EFR tone (240 Hz) at all is very impressive. The main benefit of the modified absorption silencer is the reduction of exhaust noise. However, there are also some other advantages that are highly beneficial, such as: the reduction of noise; possession of a twin wall; the property of being pre-insulated, light and portable; the property of being of a minimal length and weight; possessing an inlet and an outlet that suit modular character; being light weight; having low vibration ability; being easy to build and inexpensive - complex equipment and mounting kits are not needed. In the market, the financial criterion is of crucial importance.[14, 33]. In addition, the modified silencer is easily designed and re-assembled.

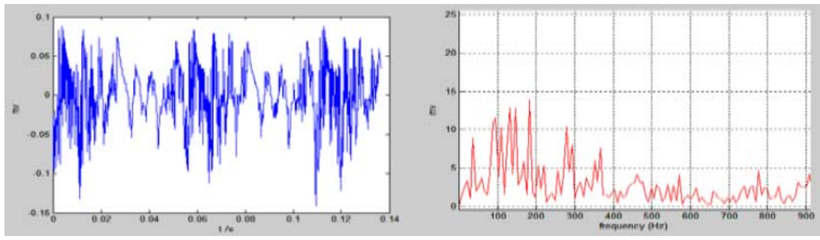


Figure 17. Time domain chart and spectrum of new absorption silencer

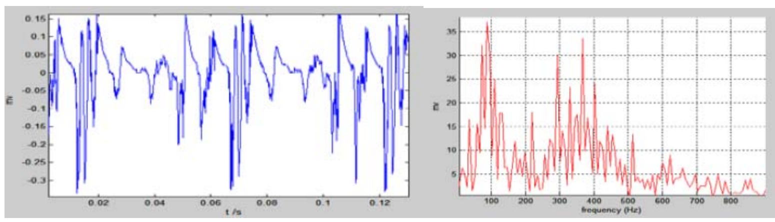


Figure 18. Time domain chart and spectrum of local or traditional silencer

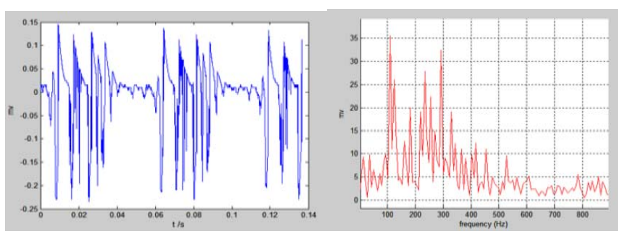


Figure 19. Time domain chart and spectrum of without silencer

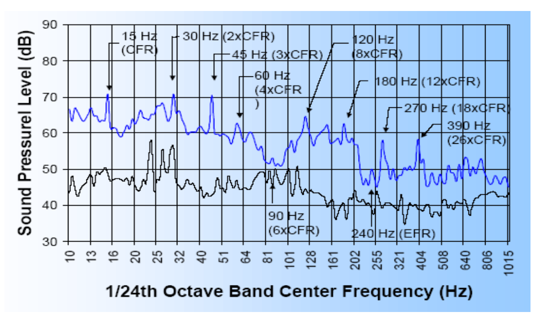


Figure 20. Sound Pressure Level of CFR & EFR with Proposed silencer

VI CONCLUSION

The experiment was performed successfully with good conditions. All the spectrums have been observed, in addition to the rules concerning its modification. This paper proposed a practical approach and the importance of a methodology to create a modified exhaust silencer. This design methodology gave a clear basic concept and will help anyone. It saves production time and cost with the easy and simple design. The experiment's conditions and the testing method are correct but the silencer was only tested witha 500KW generator set which ran at 1800RPM. It usually causesreduction of exhaust gas flow noise. Further work has to be done to test this absorption silencer with various generator sets such as 1 MW and 2MW engines. Additionally, the inclusion of transmission loss was included by using the TMM. It will developedwith the frequency range in the future in order to give a reliable expected value.

REFERRENCES

[1] G. Stewart, "Acoustic wave filters," *Physical Review,* vol. 20, p. 528, 1922.

[2] P. Davies, "The design of silencers for internal combustion engines," *Journal of Sound and Vibration,* vol. 1, pp. 185-201, 1964.

[3] J.-D. Chazot, B. Nennig, and E. Perrey-Debain, "Performances of the partition of unity finite element method for the analysis of two-dimensional interior sound fields with absorbing materials," *Journal of Sound and Vibration,* vol. 332, pp. 1918-1929, 2013.

[4] B. B. Ghosh, K. Bose, and S. Chakraborty, "STUDIES ON THE NOISE LEVEL AND PERFORMANCE OF A TWO CYLINDER DIESEL ENGINE USING MODIFIED MUFFLER."

[5] I. Dyer, "Noise attenuation of dissipative mufflers," *Noise Control,* vol. 2, pp. 50-79, 1956.

[6] D. E. Baxa, "Noise control in internal combustion engines," *JOHN WILEY & SONS, INC, 605 THIRD AVE., NEW YORK, NY 10158, 1982, 520,* 1982.

[7] D. Guicking and E. Lorenz, "An active sound absorber with porous plate," *Journal of Vibration and Acoustics,* vol. 106, pp. 389-392, 1984.

[8] F. Mechel, "Hybrider Schalldamplfer, Patent No," *DE4027511*.

[9] M. Furstoss, D. Thenail, and M.-A. Galland, "Surface impedance control for sound absorption: direct and hybrid passive/active strategies," *Journal of Sound and Vibration,* vol. 203, pp. 219-236, 1997.

[10] S. Beyene and R. A. Burdisso, "A new hybrid passive–active noise absorption system," *The Journal of the Acoustical Society of America,* vol. 101, pp. 1512-1515, 1997.

[11] P. Cobo, J. Pfretzschner, M. Cuesta, and D. K. Anthony, "Hybrid passive–active absorption using microperforated panels," *The Journal of the Acoustical Society of America,* vol. 116, pp. 2118-2125, 2004.

[12] F. B. Randall, "Industrial noise control and acoustics," *Louisiana Tech Universitz Ruston, Louisiana, USA,* 2001.

[13] D. D. Davis Jr, G. M. Stokes, D. Moore, and G. L. Stevens Jr, "Theoretical and experimental investigation of mufflers with comments on engine-exhaust muffler design," 1954.

[14] J. Igarashi and M. Toyama, "Fundamentals of Acoustical Silencers (I)," *Aeronautical Research Institute, University of Tokyo, Report,* 1958.

[15] T. MIWAT and J. IGARASHI, "Fundamentals of acoustical silencers," 1959.

[16] A. Babu, G. A. P. Rao, A. K. Sinha, and E. A. Kumar, "Simulation of a New Design Muffler to Reduce Noise in Exhaust System of C-12 SI Engine."

[17] N. Garg, O. Sharma, and S. Maji, "Experimental investigations on sound insulation through single, double & triple window glazing for traffic noise abatement," *Journal of Scientific & Industrial Research,* vol. 70, pp. 471-478, 2011.

[18] J. H. Rindel, "On the influence of low frequencies on the annoyance of noise from neighbours," *한국소음진동공학회 국제학술발표논문집,* pp. 1500-1503, 2003.

[19] L. L. Beranek and I. L. Vér, *Noise and vibration control engineering*: John Wiley & Sons New York, 1992.

[20] K. Ahuja, R. Funk, J. Hsu, M. Benne, M. L. Rivamonte, and C. Stancil, "DOT/FAA/ND-97/10," 1997.

[21] G. Jerry and P. Lilly, "Engine Exhaust Noise Control," *JGL Acoustics,* 2008.

[22] S. Shah, S. Kuppili, K. Hatti, and D. Thombare, "A Practical Approach towards Muffler Design, Development and Prototype Validation," SAE Technical Paper2010.

[23] Y.-l. Shao, "A study on exhaust muffler using a mixture of counter-phase counteract and split-gas rushing," *Procedia Engineering,* vol. 15, pp. 4409-4413, 2011.

[24] S. Gerges, R. Jordan, F. Thieme, J. Bento Coelho, and J. Arenas, "Muffler modeling by transfer matrix method and experimental verification," *Journal of the Brazilian Society of Mechanical Sciences and Engineering,* vol. 27, pp. 132-140, 2005.

[25] L. Eriksson and P. Thawani, "Theory and practice in exhaust system design," SAE Technical Paper1985.

[26] M. Munjal, "Acoustics of Ducts and Mufflers,(1987)," *Publ. by John Wiley & Sons*.

[27] I. Page, V. Page, V. Page, I. Page, X. Page, X. Page, X. Page, P. XVIII, X. Page, and X. Page, "Buyers Guide," 1997.

[28] J. G. Lilly, "Recent advances in acoustical glazing," *Sound and Vibration,* vol. 38, pp. 8-13, 2004.

[29] O. Vasile and N. Enescu, "The acoustic multi-chamber muffler performances," *Scientific Bulletin., Series D–Mechanical Engineering,* vol. 71, pp. 63-74, 2009.

[30] M. Munjal, A. G. Galaitsis, and I. L. Vér, "Passive silencers," *Noise and Vibration Control Engineering: Principles and Applications, Second Edition,* pp. 279-343, 2006.

[31] A. Selamet, F. Denia, and A. Besa, "Acoustic behavior of circular dual-chamber mufflers," *Journal of Sound and Vibration,* vol. 265, pp. 967-985, 2003.

[32] A. Graefenstein and W. Wenzel, "“Herschel-Quincke Spiral” A New Interference Silencer," SAE Technical Paper2003.

[33] Y.-C. Chang, L.-J. Yeh, M.-C. Chiu, and G.-J. Lai, "Computer aided design on single expansion muffler with extended tube under space constraints," *Tamkang Journal of Science and Engineering,* vol. 7, pp. 171-182, 2004.

Pocket Switched Networks Routing: A Survey

Rupa Barua, Shauvik Shadman, Amitabha Chakrabarty
Department of Computer Science and Engineering
BRAC University
45 Mohakhali, Dhaka-1212, Bangladesh
Email: {rupa.barua.rb, svikbrac}@gmail.com, amitabha@bracu.ac.bd

***Abstract*—Pocket switch network (PSN) is a type of delay tolerant network which is a suitable process for areas where there is no Internet connection. Such networks are also called intermittently connected networks. PSN basically works on the basis of human mobility. Here, end-to-end connectivity is not assured. There are several routing protocols available in PSN. This is a new and very attractive research domain. Since its inception PSN has seen various proposals for efficient routing in an infrastructure-less scenarios where human mobility is the only way to transfer information. We have gathered simulated results from different papers and put them together in this study to understand the concept of the different protocols and compare them in terms of delivery ratio, latency average, number of forwarded messages, and number of messages dropped.**

Keywords: DTN, PSN, routing, epidemic, spray and wait, bubble rap .

I. Introduction

We have known TCP/IP as long as Internet has been around us. Although it has been around for so long there is a drawback of this protocol. If the end to end connectivity between two nodes is somehow lost or broken, this protocol fails to ensure reliability. To overcome that, the concept of nodes carrying the message around till it finds the destination itself or a node that is close to it, was introduced (also referred to as "store-carry-forward" approach). Opportunistic Networks [31] also known as Pocket switch network (PSN) state that a human carrying a mobile phone can be both an end-user (destination) and a router (a relay node). It is one form of wireless communication network (independent of end to end connectivity between nodes) which is an instance of an older network commonly known as DTN (Delay Tolerant Network) [2]. On the other hand [20] has also categorized DTN as: flooding based, forwarding power based and social based. In flooding, messages are passed to everyone regardless of probability, whereas in forwarding a lot of metrics are taking into accord before routing. PSN is sort of network which cannot use traditional end to end connectivity or TCP/IP protocols, rather PSN uses the opportunistic meetings of human beings specifically for forwarding messages or packets whereas DTN takes into account all sorts of possible carriers including human beings, to forward data [22].It is suitable for circumstances which can lead to little or no Internet connection (for example: natural disasters, rural areas, deep forests, etc.). Since PSN is moderately a new field to explore and research, it can work without a specific infrastructure. Recent study shows that human based network connections are less inconsistent and long-termed than connections based solely on node mobility [4]. Therefore, for extreme situations social based DTN was more emphasized than any other form of DTN. For that reason PSN was introduced because it is the only form of DTN which concerns human behavior. Moreover, the current trend to opportunistically route in PSN is to base routing on human behavior and to take up a social approach which are less volatile and may lead to better routing [24]. In this paper, routing algorithms that are used in PSN are discussed and compared in terms of specific fields. All information has been taken from existing research papers, and to show our results on graphs for comparison, we generated our own data using the ONE simulator [9].

II. PSN ROUTING ALGORITHMS

This section overviews most of the popular PSN routing algorithm in detail. Although this is a new research domain, but still there are quite a number of algorithms available that requires some survey.

A. Epidemic routing

The main idea of Epidemic routing [3] has a "store-carry-forward" approach: nodes that can receive packets, store it into its buffer and carries that packet as it moves, passing the packet along to new nodes as they come into contact. Whenever the packet-carrying (source) node comes across a node that does not have a copy of that packet already, the source node is said to "infect" this new node by passing the copy along and the newly infected node behaves the same when it comes into contact with other susceptible nodes (i.e., nodes without a packet).This routing protocol trades off performance by achieving minimum delivery delay with an increased usage of resources like transmission power, buffer size, bandwidth etc. There are some recovery schemes associated with this protocol [1]. Firstly, after the packet has been delivered to the destination node, a node can generate an "anti-packet" within itself so that others nodes would not pass along the same packet again. This is known as the "IMMUNE" recovery scheme. A more strict approach according to [1] is the forwarding of the "anti-packet" among the infected nodes (which is known as the "IMMUNE-TX" scheme) or among both the infected and susceptible nodes (which is known as the "VACCINE" scheme) so that the number of copies sent are reduced. Both the "IMMUNE-TX" and "VACCINE" have similar buffer requirements.

B. First-Contact

The concept of first contact routing [4] involves only a single copy of the message available in the whole network. A node forwards the packet only when any single contact is available. If none of the contacts or paths are available, the message waits for one to be available. The source node passes along the copy of the message the first node it comes in contact with, making that node a relay node if it is not the destination itself. Once the message is passed along, the node deletes that message from its buffer. To make sure two nodes did not exchange the same message back and forth, a node forwards the message only to nodes who did not have it at all. Then after passing along and deleting the copy from its buffer it generates the "anti-packet" so that it does not get re-infected. This results in having a bad-delivery ratio, because the next node is selected randomly, which does not guarantee that this node has a higher probability of contacting the destination node than the previous, so no high yield. Moreover, even if the previous node had a greater chance to reach the destination node, it cannot be re-infected. This routing protocol only works if the source and the destination is only one hop away.

C. Bubble-rap

What is a community? It has been a vital concept of sociology and ecology for a long period of time. Community is a term used to assemble people who have common taste or maybe living in the same location [5]. This protocol is solely based on social behavior of humans. As in, it operates following a trend of popularity (connectivity). All the nodes in the network are grouped into a community and the node passes along a message based on the popularity 'RANK', usually if the next node is in a higher rank than the current one [4]. For this protocol to work, every node must belong to a community and have two rankings: local and global. The global ranking labels the node in the entire society whereas local ranking denotes the node's place in its own community [5]. If the destination node is within the community then the message forwarding will depend on a higher local ranking than the current node. Otherwise, the forwarding of the message will depend on a higher global ranking until it comes in contact with a node in the destination's community. From then on, higher local ranking is used to forward the message. By doing so, the probability of reaching the destination node will be greater. But what if one person belongs to multiple communities, that is, what if the communities overlap? It is vital to detect this feature. The K-clique method completes this purpose, and it is designed for binary graphs, so specifying the threshold of the edges is important. WNA (Weighted Network Analysis) can work on weighted graphs directly but cannot detect overlapping communities. We use both as per our needs.

D. Spray and wait

Spray and wait is another routing protocol for Pocket switched networks which was proposed in [10]. It is a modified

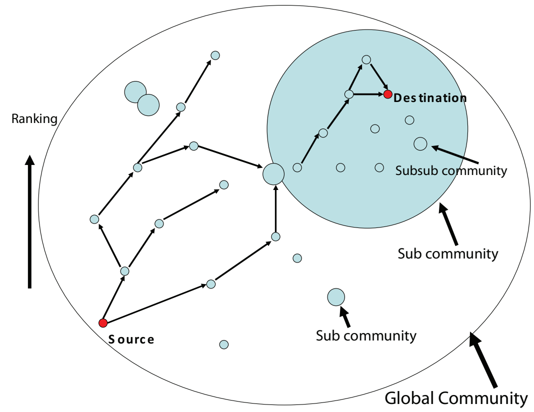


Fig. 1: Overall view of how the bubble rap works. [4].

version of the flooding based protocol. The first definition of Spray and wait is that it has two phases, which are as follows:

1. Spray phase: every message that the source node carries, L number of copies of the message are initially forwarded by the source node and possibly other nodes receiving a copy - to L number of different "relays".

2. Wait phase: if the destination is not found in the spraying phase, each of the L nodes carrying a copy of the message forwards it only to its destination (performs direct transmission).

Epidemic routing and flooding has been morphed, resulting to Spray and Wait. Flooding keeps giving out copies of the message to every node it encounters until it reached its destination through nodes who received the copy or through directly passing. But Spray and wait makes L copies and sends them to L distinct nodes. Those nodes keep the copy until they meet the receiver. Here the number of message copies and how many to share remains open to discussion. In spray and wait, if there are L number of copies then any node A that has $n > 1$ message copies (source or relay), and encounters another node B (with no copies), it hands over half the amount $(n/2)$ to B and keeps half for itself; when it is left with only one copy, it switches to direct transmission.

E. Lobby Influence

Lobby Influence was proposed in [6]. Lobby Influence works similar to Bubble Rap in the sense that it uses opportunistic ways to pass messages and they are both social based forwarding algorithm. It uses a modified version of bubble rap to enhance the delivery ratio and put less stress on the most popular node. The basis for Lobby Influence is derived from [7] who presented the metric known as Lobby Influence. They used diplomats dilemma [8] which states that a diplomat has a high influence in the society because he knows a lot of influential people of the society. Therefore a diplomat is as important as the actual influential person and he can reach them with minimum effort and low cost. The culmination of two criteria gives rise to the Lobby influence routing protocol. The two criteria are: Node Popularity (np) and Lobby Index (li). In summary Lobby Influence takes up

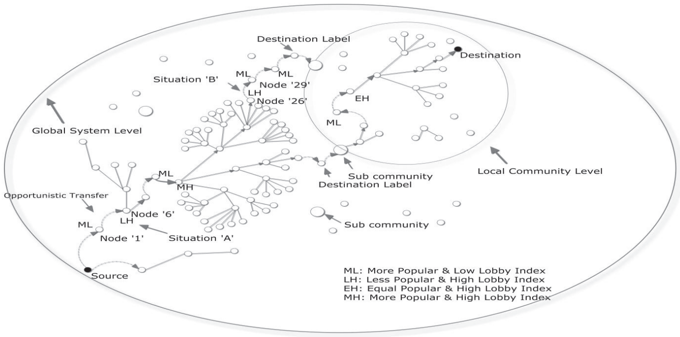


Fig. 2: Working principle of lobby influence. [6]

the ideas of Bubble Rap and Lobby Index algorithms and merges them together which results in the overcoming of the shortcomings of both the algorithms. The algorithm proposed by Khan et al may come across two situations on the basis of their algorithm:
1. Node Within a local community: In this situation if the destination node is in the local community and if the encountered node is part of the local community then local rank and lobby index is used to determine the forwarding decision. If lobby index or rank is higher than it forwards.
2. Node within the global system: In this situation the destination node is part of the global community. If the encountered node is part of the same community as the destination node then the message will be transferred and deleted from the node which it was sent from.

Khan et al simulated this algorithm using ONE (Opportunistic Networking Environment) [9] simulator which is designed for delay tolerant networks. In terms of message delivery and speed Lobby Influence outperformed Bubble Rap and Epidemic routing algorithms. According to [6], LI has a higher communication cost compared to BR but it significantly reduces load on most popular nodes in the network.

F. Prophet

Prophet algorithm is proposed in [13] which stands for Probabilistic Routing Protocol using History of Encounters and Transitivity. They have hugely relied on the repeated behavioral pattern which means that if a node visits a location several times, it is likely to visit that location again. Delivery Predictability is a probabilistic metric which was used by Lindgren et al which is defined as $P(a, b)$ at every node a, for each known destination b. This is used to understand the level of probability node a has to deliver messages to node b. Prophet is similar to epidemic routing in the sense that when two nodes meet they exchange messages as well as delivery predictability information stored at the nodes.They have calculated the delivery predictability of messages which is briefly described below:

Firstly the metric update is taken into consideration. The metric needs to be updated every time a node is encountered so that it is understood which nodes have a high delivery predictability. The following calculation has been used:

$$P_{a,b} = P_{(a,b)old} + (1 - P_{(a,b)old}) \times P_{init} \tag{1}$$

Secondly, the delivery predictability must age because if one node does not meet another certain node for a while then according to their hypothesis it is less likely that they will meet again. An aging constant, has been used in the following equation:

$$P_{a,b} = P_{(a,b)old} \times \gamma^k \tag{2}$$

Thirdly, the delivery predictability has a transitive property. This basically means that if node A meets node B frequently and node B meets node C frequently then it is highly probable that a message given to C will be delivered to A. They have used a scaling constant in the following equation:

$$P_{a,c} = P_{(a,c)old} + (1 + P_{(a,c)old} \times P_{a,c} \times P_{b,c}) \times \beta \tag{3}$$

Finally the forwarding strategy is pretty simple and straight forward. When two nodes meet if the delivery predictability of the destination of the message is higher at the encountered node then the message is passed.

G. Friendship-based routing algorithm

In PSN, since the people are considered as nodes, this algorithm involves making the decision based on the friendship between two nodes (people). This algorithm introduced a new metric called social pressure metric (SPM), taking into account different sides of social behavior of people. For two people to be considered friends of each other, they have to meet up frequently, make regular and long-lasting contact. In order to ease up the challenges faced during discontinuous end-to-end connectivity, [14] have emphasized on three components of friendship: durability, high frequency and regularity. Friendships can strong and weak. Two nodes can be good friends directly, other scenarios include two nodes having no direct friendship among them but has a very strong mutual friend. In that case, they can be considered as indirect close friends. To label such indirect connections, [14] suggests to use conditional SPM between. And when it comes to direct friendships, every node can identify them using their own contact history.

This algorithm follows the following forwarding strategy: if node A has to forward a message to node B, and meets node C in the middle, A will forward the packet to C if and only if C has a stronger friendship with B than A in the current time period. But in [14], it was also mentioned that even if C has a better connection to B than A but does not include B in its current time period (time taken to form a community), A will not forward it to C.

III. Result Comparisons

This section compares results of various routing algorithms for PSN. The comparisons are based on simulations and comparisons from other research papers. The references will be mentioned accordingly.

A. Prophet vs Epidemic:

To start with, we can see according to Lindgren [13], as the queue size increases the average delivery rates of PROPHET is always above the curve for epidemic routing.

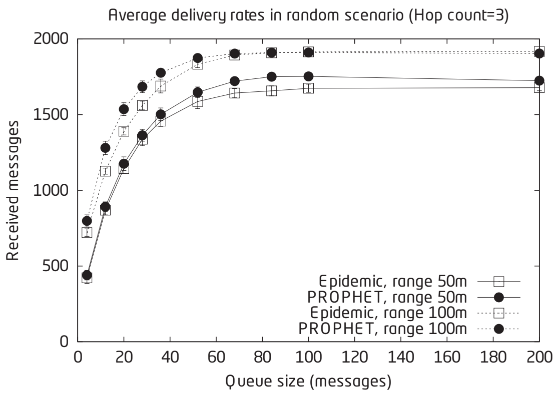


Fig. 3: Average Delivery Rate in random scenario (Hop Count=3). [13]

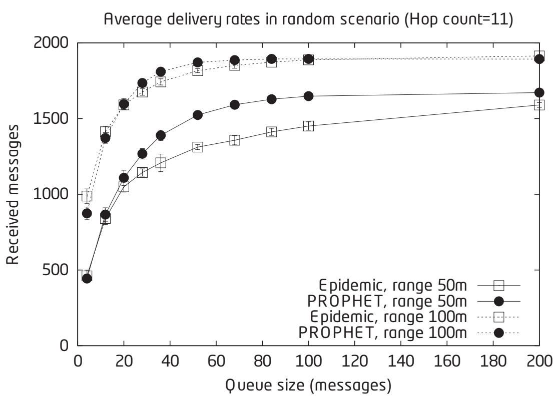


Fig. 4: Average Delivery Rate in random scenario (Hop Count=11) [13]

Secondly, we have the graphs for the average delays in random scenario where upto a certain queue size the delay in PROPHET is more than Epidemic but as the queue size increases, Epidemic tends to have higher delays than PROPHET.

Also in fig7 and 8 the average delivery rates (in community scenario) is shown as a comparison with Epidemic. It can be seen that by all accounts as the queue size increases the number of received messages is always more for PROPHET by a significant amount. Hence, we can conclude by saying that PROPHET is better in delivery rates (both in random scenario and community scenario), and also having lower delays when in comparison to Epidemic Routing.

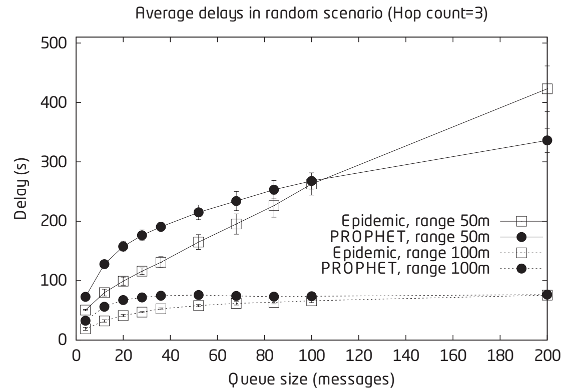


Fig. 5: Average Delays in random scenario (Hop count=3). [13]

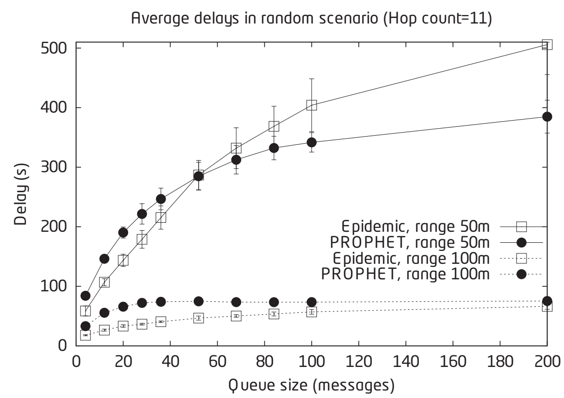


Fig. 6: Average Delays in random scenario (Hop count=11). [13]

B. Lobby Influence vs Epidemic vs Bubble Rap

This experiment was done in [6]. They used a total of ten experiments on 24 Computers. The code for this comparisons is provided in [9]. The algorithms were measured against three metrics: 1. Message Delivery: How many packets arrive at destination 2. Delays: How quickly packets arrive at the destination 3. Forwarded messages: the cost in terms of number of exchanged messages, in other words utilization of resources.

1. Message delivery: In fig9 it can be seen that Lobby influence has the highest received messages as the queue size increases at any given point during the course of the experiment. Epidemic on the other hand performs worse than Bubble rap.

2. Delays: The latency average is measured in seconds in fig10. Epidemic has the highest latency because most of the nodes exhaust their queue size by accepting unnecessary messages due to lack of message forwarding criteria and as a result more messages are dropped. Here we see that Lobby Influence has the best Latency. This means that packets or messages are being delivered faster than BR and Epidemic.

3. Forwarded messages: The amount of messages forwarded

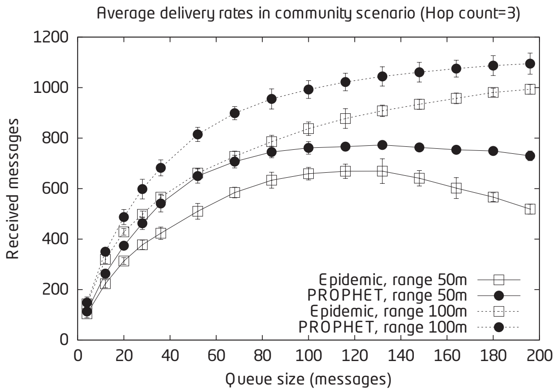


Fig. 7: Average Delivery Rate in random scenario (Hop Count=3) [13]

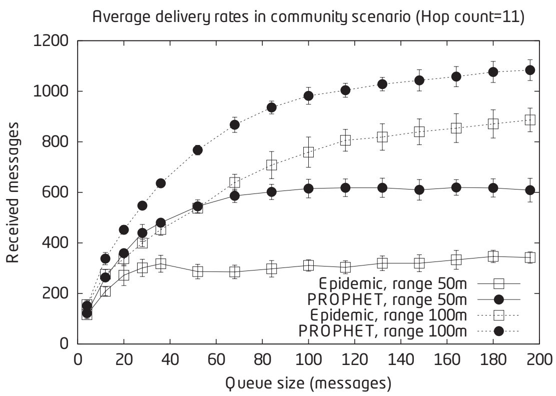


Fig. 8: Average Delay Rate in random scenario (Hop Count=11) [13]

by Epidemic is rightfully at a higher level than the other algorithms because of the way epidemic works. Epidemic basically gives those messages, the encountered node does not have, to every node it meets. Therefore it passes messages very frequently when compared to the others. Bubble rap only gives messages based on one criteria: rankings. On the other hand Lobby Influence gives the messages according to two criteria's: rankings and Lobby influence so probable message sharing is more. Although bubble rap has the best figures here it can be said that due to the delivery ratio and lower delay lobby influence is better.

C. Spray and wait vs other algorithms:

Spyropoulos et al proposed the spray and wait and simulated this using a custom event-driven simulator and evaluated them using a variety of mobility models and under contention. In this scenario a network size of $200x200$ has been used. The number of nodes M and the transmission range K has been varied to compare the performances of several routing algorithms. A lot of combinations have been used from very sparse, highly disconnected networks to very closely spaced networks. In this simulation the following algorithms have been considered: (1) Epidemic routing (2) Randomized flooding (3) Utility-based

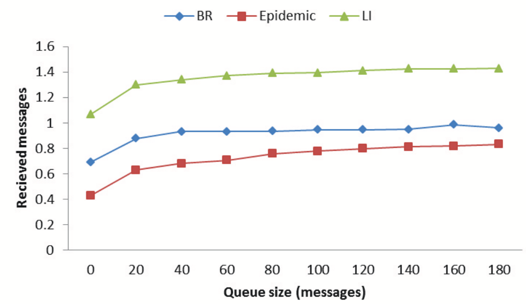


Fig. 9: Average Received Messages in Bubble Rap, Epidemic and Lobby Influence. [6]

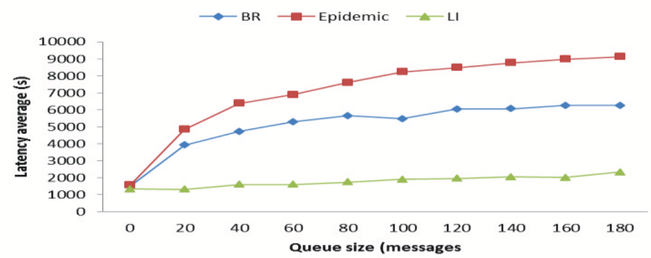


Fig. 10: Latency average in Bubble Rap, Epidemic and Lobby Influence. [6]

routing (4) Spray and Wait (5) Seek and Focus single-copy routing (6) Oracle-based Optimal routing

In figure12 and13 the number of transmissions have been tested. Here, the less the number of transactions, the better the algorithm. That means, that the less the number of transmissions the less amount of resources have been used of a single node. It is evident from the graph that with varied transmission range K values and $M = 100$ and $M = 200$ that spray and wait has the smallest transmissions and epidemic has the highest number of transmissions. Therefore we can safely assume that spray and wait in comparison to the mentioned algorithms is better at resource utilization.

In scenario fig14 and 15, the delivery delay has been monitored using different values for K and M. It can be observed that spray and wait (both versions where $L = 10, 16, 20, 32$), has the lowest delivery delay among the mentioned algorithms.

Therefore considering all the aspects we can say that comparing Spray and wait with Epidemic, it is better in all aspects like delivery delay and number of transmissions.

The results have been summarized in Table I.

One of the challenging issues in PSN is network congestion (due to flooding). In order to obtain improved results in this area, we can look into how the nodes take forwarding decisions. In [32], they talk about a heuristic function based on the hop count, that helps the message-carrying nodes take forwarding decisions. That is done by using the information carried by the packets traveling through the network. Other

TABLE I: Comparison between different Algorithms .

	Epidemic	First-Contact	Spray and wait	Prophet	Lobby Influence	Bubble Rap
Messages delivered	Moderately Low	Very low	Low	High	Very High	Moderately High
Messages dropped	Very High	Very Low	Moderately High	High	Moderately Low	Low
Messages forwarded	Poor	Excellent	Fair	Adequate	Good	Very Good
Average Latency	Fair	Poor	Good	Adequate	Very Good	Excellent

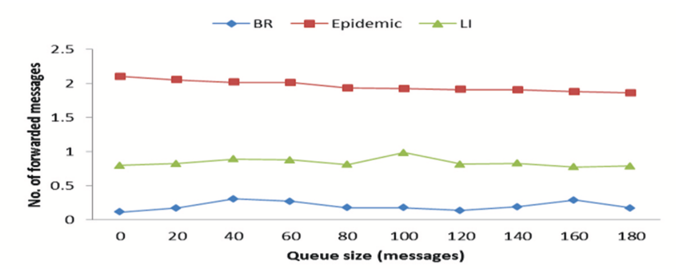


Fig. 11: Number of forwarded messages in Bubble Rap, Epidemic and Lobby Influence. [6]

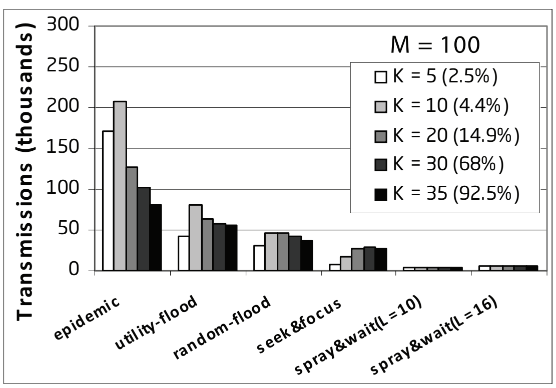


Fig. 12: Transmissions as a function of number of nodes M=100 and transmission range K. [1]

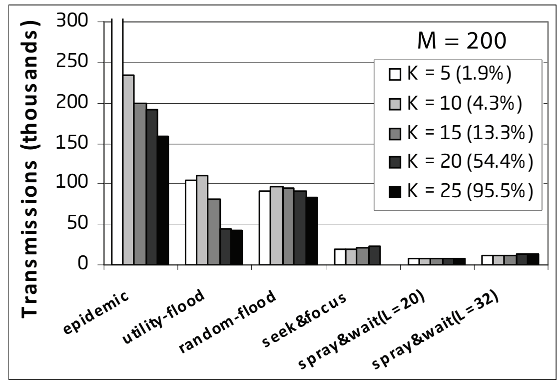


Fig. 13: Transmissions as a function of number of nodes M=200 and transmission range K. [1]

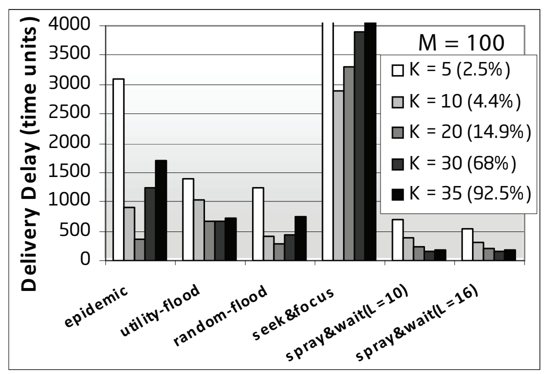


Fig. 14: Delivery Delay as a function of number of nodes M=100 and transmission range K. [1]

forwarding schemes include the direction entropy-based forwarding scheme (DEFS) [33], which involve emphasizing only on those nodes that are more prone to travel to different locations to forward the messages to the destination nodes. In addition to this, since the nodes are always moving, the topology of the whole network is a bit difficult to keep track of. A node will always have to either search for the destination, decide to stop or send a response back. Even if we do succeed in keeping track, loss of resources (like battery) has to be taken into account. In [34], an optimal search was suggested where there are three parts: static search in which the search depth is set at the start of a query, dynamic search in which the depth is determined locally during the forwarding of messages, and learning dynamic search which influences the observation to determine whether the content is suitable for the query or not.

IV. Future Research Scope:

For 2017 the number of mobile phone users is forecasted to reach 4.77 billion. So we can assume that PSN will be very useful and popular in the near future. If we can implement PSN using all the 4.77 billion cell phones we can have almost an infallible network. There is still a lot of work that needs to be done to implement pocket switched networks. Work on reliability of the messages and to control the huge amount of data (which can lead to congestion in the network) is done in [19] but they are based on MANETS, which can easily be implemented on PSNs. In the near future, space programs may need to implement PSNs which is still open for research as mentioned in [21]. Work is done in [27], [28], [29] showing that Wi-Fi connections of the cell phones may be used to

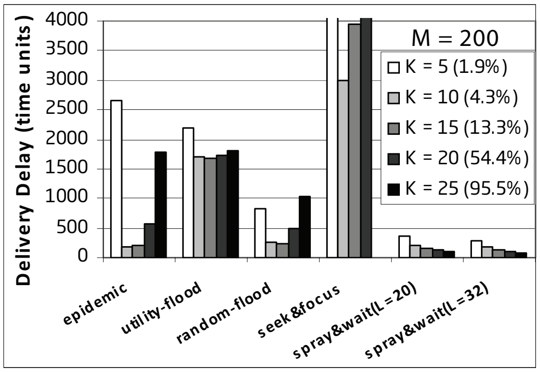


Fig. 15: Delivery Delay as a function of number of nodes M=200 and transmission range K. [1]

create a pocket switched network to send and receive messages from a distant location. There is still room for improvement in this and the routing protocols for better delivery ratio and lower drop rate. A very big issue in PSN is: there is very little or almost no work done in the security measures. One of the attacks is known as flood attacks which is an issue for DTNs mostly, but may also be an issue in PSNs. In [23] they address this issue by implementing rate limiting [30] but, then again, this is for DTNs. The basic idea is very simple, which is to limit the source node to transmitting a ceiling value of messages. Therefore in the future, the security of PSN may still be researched upon to create a full-proof system of passing messages.

V. Application Domain:

The very first application domains that can be noticed are the areas where there are no infrastructures for any end to end connectivity. PSN should be normally applied using cell phones. For example, in a dense forest this can be used since very few people travel through there. If everyone has PSN enabled cell phones then data collection and sharing of data would be very easy (since no connection is required). The people living in very distant rural areas may use PSNs as non-interactive internet [25]. Secondly the work of [15] on the reindeer herd of the Swedish Lapland's is mentionable. They have used DTNs, but they could also use PSN which would perform the same. Thirdly, NASA has a lot of satellites in space and using that we get weather information, short/long range environmental prediction, global air current prediction, and also predicting natural hazards according to [16].A way has been searched such that satellites can share information. In their paper, they have discussed the sharing of such data for earth observing satellites that will support the next generation of space exploration. In [21] they suggest that satellites, which can be subjected to tolerate delays and where long space and satellite communications is needed, we can use DTN. Moreover, they have pointed out that DTN is the way to go to support future space programs and deep space communication which is also supported by [25], [26]. Again, since PSN is a part of DTN this can also be applied here. Another application domain is collecting information on animal behavior, ecology, habitat preference of animals, physiology and movement patterns [17]. The use of PSN may greatly improve efficiency. We can use the work done at [18] to know the weather information throughout a park. The military may also use PSN specifically to gather information and to transmit messages about their environment among their own respective sides. Since in a war zone there is absolutely no internet and no cellular networks they may use PSNs [25], [26]. Lastly, in extreme situations like natural disasters, we can use PSN in rescue missions. In a disaster recovery scene we can use PSN to find victims and gather vital information about the rescue workers and also to station them in an efficient manner [21], [25], [26].

VI. Conclusion

In this paper we have provided a brief literature review of several routing protocols that are applied on pocket switched networks, which is a sub division of Delay Tolerant Networks. Through our observation of the literature we have come up with the conclusion that Lobby Influence outperforms all the other routing protocols in terms of delay time, the drop rate, the number of messages delivered and the number of forwarded messages. The results of several simulations have been mentioned and it is evident that among the many famous routing protocols like epidemic, bubble, PROPHET that the Lobby Influence routing protocol performs the best. But there are still works that can be done here to improve the delivery ratio and the delay. The application domain of Pocket Switched Networks have also been suggested and we have also suggested some areas of PSN which are still open for research and can progress this field further.

In a world with 4.77 billion cell phones roaming around, imagine each cell phone turned into a mobile router. Internet, as we know it, will change forever. Even if every static router on earth goes down we will still be able to get messages to and from others throughout this huge network. We will never be disconnected.

References

[1] Hossmann, T., Spyropoulos, T., Legendre, F. *Putting contacts into context: Mobility modeling beyondinter-contact times.* In Proceedings of the Twelfth ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc '11).

[2] Sarkar, R., Rasul, K., Chakrabarty, A.*Survey on routing in pocket switched network.* Wireless Sensor Network. (2015).

[3] Zhang, X., Neglia, G., Kurose, J., Towsley, D.*Performance modeling of epidemic routing.* Computer Networks (2007).

[4] Gamit, V., Patel, M.*Evaluation of DTN Routing Protocols.* International Journal of Engg. Sciences and Research Technology.

[5] Hui, P., Crowcroft, J., Yoneki, E.*Bubble rap.*Proceedings of the 9th ACM International Symposium on Mobile Ad Hoc Networking and Computing (Mobihoc '08).

[6] Khan, S., Mondragon, R., Tokarchuk, L.*Lobby Influence: Opportunistic forwarding algorithm based on human social relationship patterns.* IEEE International Conference on Pervasive Computing and Communications Workshops (2012).

[7] Korn, A., Schubert, A., Telcs, A.*Lobby index in networks* Physica A: Statistical Mechanics and its Applications (2009).

[8] Holme, P., Ghoshal , G.*The diplomat's dilemma: Maximal power for minimal effort in social networks* In Adaptive Networks n Adaptive Networks : Springer Berlin Heidelberg (2009).

[9] Dillon, P. *Software contribution to the ONE simulator [Online]* Available: www.cs.pitt.edu/ pdillon/one

[10] Mangrulkar, R., Atique, M. *Performance Evaluation of Flooding Based Delay Tolerant Routing Protocols* In Proceedings of the National Conference on Emerging Trends in Computer Science and Information Technology (February ’11).

[11] Spyropoulos, T., Psounis, K., Raghavendra, C. *Spray and wait: an efficient routing scheme for intermittently connected mobile networks.* Proceedings of the 2005 ACM SIGCOMM workshop on Delay-tolerant networking (ACM ’15).

[12] Kernen, A., Ott, J., Krkkinen, T. *The ONE simulator for DTN protocol evaluation.* Proceedings of the 2nd international conference on simulation tools and techniques (ICST ’09).

[13] Lindgren, A., Doria, A., Scheln, O. *Probabilistic routing in intermittently connected networks.* ACM SIGMOBILE mobile computing and communications review (ACM ’03).

[14] Bulut, E., Szymanski, B. K. *Friendship based routing in delay tolerant mobile social networks.* Global Telecommunications Conference (GLOBECOM 2010).

[15] Doria, A., Uden, M., Pandey, D. *Providing connectivity to the saami nomadic community.* Generations (2009).

[16] Prescott, G., Smith, S., Moe, K. *Real-time information system technology challenges for NASA’s earth science enterprise.* Proceedings of The 20th IEEE Real-Time Systems Symposium, Phoenix, Arizona (1999).

[17] Boehlert, G., Costa, D., Crocker, D., Green, P., O’Brien, T., Levitus, S., Le Boeuf, B. *Autonomous pinniped environmental samplers: using instrumented animals as oceanographic data collectors.* Journal of atmospheric and oceanic technology, 18(11).

[18] Beaufour, A., Leopold, M., Bonnet, P. *Smart-tag based data dissemination.* Proceedings of the 1st ACM international workshop on Wireless sensor networks and applications (ACM ’02).

[19] Nagmote, P., Hande, K. *Data Replication To Increase The Reliability And To Control Congestion In Opportunistic Network.* International Journal of Advanced Research in Science, Engineering and Technology (July, 2015).

[20] Patel, C., Gondaliya, N. *Enhancement of Social based Routing Protocol in Delay Tolerant Networks* International Journal of Computer Applications (July, 2015).

[21] Kumari,S., Yadav, P., Yadav, M. *Review of Efficient Routing In Delay Tolerant Network* International Journal Of Engineering And Computer Science ISSN: 2319-7242 Volume 4 Issue 12 (December, 2015).

[22] Ali, M., Fida, M., Arsalaan, A., Adnan, A. *Implementation and Analysis of Clustering Techniques Applied on Pocket Switched Network* International Journal of Distributed Sensor Networks Volume 2015, Article ID 239591 (September, 2015).

[23] Niharika, M., Sivaji, U. *Defending Against Flooding Attacks in Mobile Networks* International Journal Of Scientific Engineering and Technology Research Vol.04,Issue.15(June,2015).

[24] Basith, A. *A Framework for Routing Performance Analysis in Packet Delivery Delay Tolerant Networks with Application to Non Cooperative Networks* International Journal of Computer Science information and Engg., Technologies (2015).

[25] Gangele, S., Bodne, N. *ADAPTIVE MULTICAST FOR DTN FOR REDUCING DELAY* Global Journal of Engineering Science and Research Management (April, 2015).

[26] Wang, E., Yang, Y., Wu, J. *Dynamic Beaconing Control in Energy-Constrained Delay Tolerant Networks* Communications (ICC), 2015 IEEE International Conference on. IEEE. (2015).

[27] Hu, S., Lui, H., Su, L. Wang, H., Abdelzaher, T., Hui, P., Zheng, W., Xie, Z., Stankovic, J. *Towards Automatic Phone-to-Phone Communication for Vehicular Networking App* INFOCOM, 2014 Proceedings IEEE. (2014).

[28] Sharma, A., Navda, V., Ramjee, R., Padmanabhan, V., Belding, E., *Cool-tether: Energy efficient on-the-fly wifi hot-spots using mobile phones* Proc. of ACM CoNEXT (2010).

[29] Wang,E., Yang, Y., Wu, J. *Energy Efficient Phone-to-Phone Communication Based on WiFi Hotspot in PSN* Computer Communication and Networks (ICCCN), 2015 24th International Conference on. IEEE. (2015).

[30] Li, Q., Zhu,S. *To lie or to Comply: Defending against Flood Attacks in Disruption Tolerant Networks.* IEEE Transactions on Dependable and Secure Computing, Vol 10, No. 3, pp 168-182. (2013).

[31] Boldrini, C., Lee, K., Ott, J., Pagani, E. *Opportunistic Networks* Computer Communications (July, 2014).

[32] You, L., Li, J.,Wei, C., Dai, C.,Xu,J.,Hu,L. *A Hop Count Based Heuristic Routing Protocol for Mobile Delay Tolerant Networks* Scientific World Journal Volume 2014, Article ID 603547 (June, 2014).

[33] Jeon, M., Kim, S., Yoon, J., Lee, J., Yang, S. *A Direction Entropy-Based Forwarding Scheme in an Opportunistic Network* Journal of Computing Science and Engineering, Vol. 8, No. 3. (September 2014).

[34] Hyyti, E., Bayhan, S., Ott, J., Kangasharju, J.*Searching a Needle in (Linear) Opportunistic Networks* Proceedings of the 17th ACM international conference on Modeling, analysis and simulation of wireless and mobile systems. ACM. (2014).

Performance Analysis of Routing Protocols in Mobile Ad-hoc Network (MANET)

Md. Zulfikar Alom[1], Tapan Kumar Godder[2], Mohammad NayeemMorshed[3], *Student Member, IEEE*

[1,2]Department of Information & Communication Engineering, Islamic University, Kushtia-7003, Bangladesh
[3]School of Computer and Communication Engineering, University Malaysia Perlis (UniMAP), 02600Arau, Perlis, Malaysia

***Abstract*— Mobile Ad hoc Network (MANET) consists of a collection of wireless mobile nodes that are capable of communicating with each other without the use of a network infrastructure. The main procedure for data transmission in ad hoc network is routing. In this paper, we evaluate the performance characteristics and behavior of few ad hoc communicating protocols namely Ad Hoc On Demand Routing Vector (AODV), Dynamic Source Routing (DSR) and Destination Sequenced Distance Vector (DSDV) Routing Protocols in MANET. Performance comparison is based on performance metrics such as throughput, end-to-end delay and normalized routing load are evaluated using network simulator (NS-2.35) varying number of nodes and number of pause time.**

Keywords—Mobile Ad Hoc Networks (MANETs), Ad Hoc On Demand Routing Vector (AODV), Dynamic Source Routing (DSR), Destination Sequenced Distance Vector (DSDV).

I. Introduction

Mobile ad hoc mobile networks are structures where nodes communicate among each other without whatever existing commercial infrastructure and wirelessly. They have the luxury of rapid deployment, robustness, suppleness and untouched support just for mobility. This particular flexibility associated with self setting up and self administration causes it to be lucrative with regard to various programs in army operations, cellular mesh systems; wireless sensor systems etc. Because of the wireless character of Mobile ad hoc network, the redirecting protocol is really a very essential issue to create it better and dependable. Quite a few papers provide comparison between routing methodologies DSR, AODV, DSDV in addition to TORA depend on PDF, Average End to end Delay in addition to NRL [1][2]. This paper aims to produce a comprehensive comparative research of three favorite routing protocols: AODV, DSR and DSDV in MANET and in the last the conclusion will be presented, that which routing protocol is the best one for mobile ad hoc networks.

II. Routing Protocols In Ad-hoc Networks

Ad hoc network routing protocols are classified into three major categories based on the routing information updated mechanism as shown in Figure (1). There are Proactive (table driven routing protocols), Reactive (on-demand routing protocols) and Hybrid routing protocols being combination of Proactive and Reactive [1].

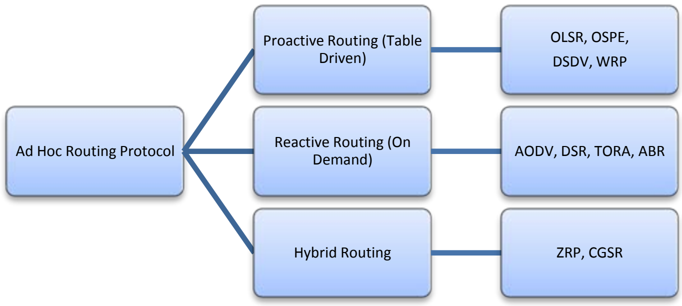


Fig.1 Classification of Mobile Ad-Hoc Routing Protocol.

A. *AODV (Ad Hoc on Demand Routing Vector)*

The ad hoc on demand distance vector (AODV) is dependent on distance vector routing algorithm. Nevertheless, unlike range vector, it's a reactive process; it demands the path when required. It doesn't require nodes which maintain paths for locations, which aren't actively utilized in communication. The options that come with AODV redirecting protocol tend to be loop-free redirecting and instant notification will be sent towards the affected nodes upon link damage [3]. The formula uses numerous messages to keep and uncover links. They are route request (RREQ), route reply (RREP), as well as route error (RERR) [4]. Whenever a source node wants to establish the communication program, it triggers a path- break through process. The origin node wide casts the RREQ packet using its IP address, broadcast ID (BrID) as well as sequence amounts of source as well as destination. As the BrID as well as IP address can be used to distinctively identify every request. Receiving node arranged the backward pointer towards the source as well as generates the RREP packet if it's the location.

B. *DSR (Dynamic Source Routing)*

DSR will allow the network for being completely self-setting up and self-configuring, without the need for almost any existing network infrastructure or administration. The protocol consists of the a couple main parts of “Route Discovery” in addition to “Route Maintenance”, which band together to make it possible for nodes to get and retain routes to help destinations from the ad hoc network [5]. A selling point of DSR is usually that nodes can certainly store many routes into their route cache, which shows that the supplier node can certainly check it is route cache for just a valid way before initiating route discovery in case a logical route is found there no requirement for way discovery.

C. *DSDV (Destination Sequenced Distance Vector)*

DSDV is mostly a table made routing protocol this really is an upgraded version for the distributed Bellman-Ford formula. In the entire table made protocols any node says a table made from the so next hop to arrive at all countries. To maintain ones tables new they really are exchanged approximately neighboring nodes located at regular intervals or every significant topology transformations are recognized [6].

III. Simulation Metrics And Parameters

The simulations were performed using Network Simulator (Ns-2), which is popularly used for ad hoc networking community. The routing protocols were compared based on the following three performance metrics:

Throughput: Throughput is defined as the rate of the total data reaches a receiver from the sender. The time it takes by the receiver to receive the last message is called as throughput. Throughput is expressed as bytes or bits per sec (byte/sec or bit/sec) [7].

End to End delay: The average time from the beginning of a packet transmission at a source node until packet delivery to a destination. This includes delays caused by buffering of data packets during route discovery, queuing at the interface queue, retransmission delays at the MAC, and propagation and transfer times. Calculate the send(S) time (t) and receive (R) time (t) and average it [8].

Normalized Routing Load: The number of routing packets transmitted per data packet delivered at the destination. Each hop wise transmission of a routing packet is counted as one transmission [9].

Routing Load = Routing Packets Sent / Received Packets

The simulation parameters are as follows:

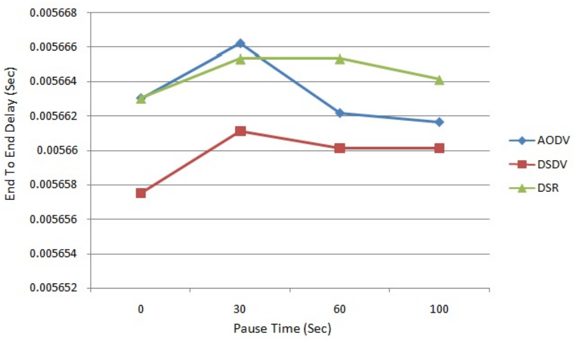
Table I: Simulation Parameters

Parameters	**Value**
Simulator	NS-2 (Version 2.35)
Channel Type	Wireless Channel
Radio Propagation Model	Two ray round wave
Network Interface Type	Wireless
Link Layer Type	Logical Link
Antenna	Omni Antenna
Simulation Area (m*m)	500*500
Simulation Duration	100 seconds
Number of mobile nodes	2,4,8,12,16
Maximum Packet Size	500
Traffic Type	CBR
Source Type	UDP, TCP
Protocol	AODV,DSR and DSDV
MAC Type	Mac 802.11
Physical Type	PHY 802.11b

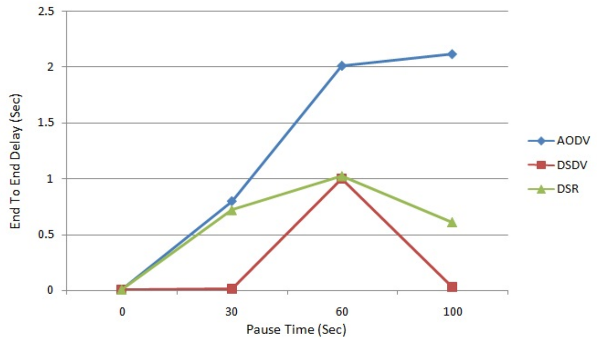
IV. Simulation Result And Analysis

A. *Varying Both Number of Nodes and Simulation Times*

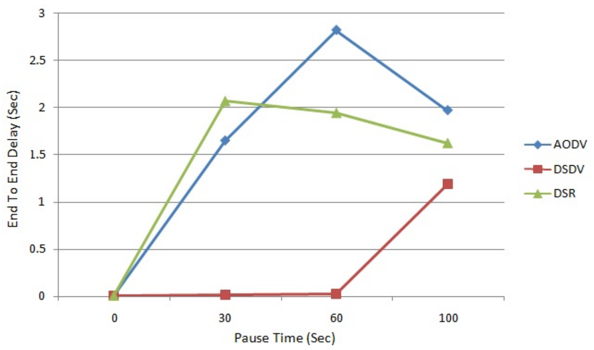
The first set of experiments uses differing the number of nodes and changing the simulations. For the 4 nodes, 8 nodes and 16 nodes experiments respectively.



2.a



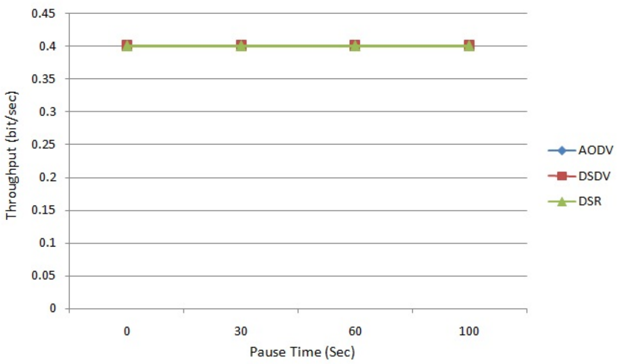
2.b



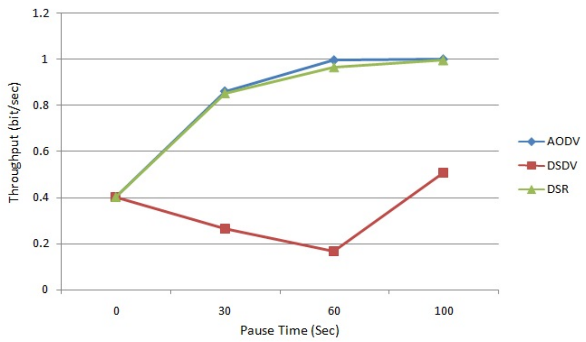
2.c

Fig:2(a,b,c) End to end delay for 4,8 and 16 node respectively with varying Simulation Time

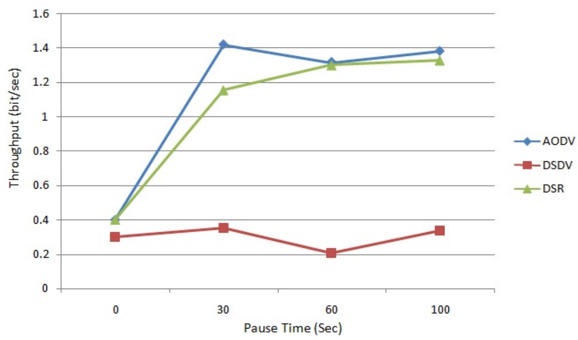
In Figure 2(a,b,c) it can be seen from the results, end to end delay is higher in AODV followed by DSR and DSDV having the lowest and most stable End to End Delay in mobility. AODV has only one route per destination in the routing table, which is constantly updated based on sequence number and DSDV has to continuously update the whole routing table periodically when needed, which leads to a slight delay in delivery. The end to end delay does not change with increase in the number of nodes as the source and destination are in the same place moving with same speed, the increased number of nodes only might increase number of hops. The End to End delay decreases with increase with speed, as when it moves more frequently the routing updates are exchanged more frequently and faster it reaches the destination.



3.a



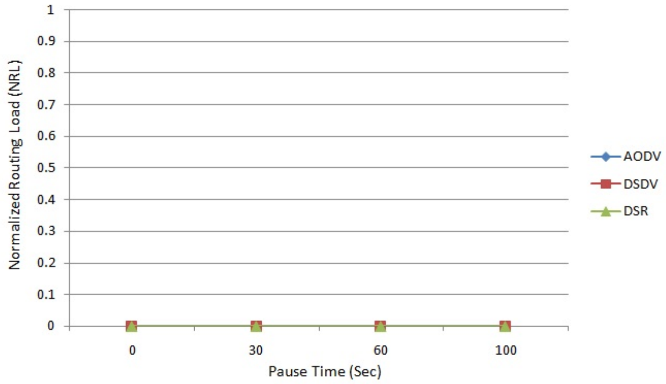
3.b



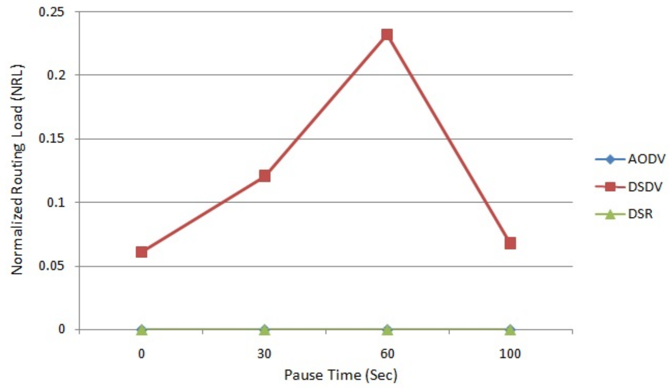
3.c

Fig:3(a,b,c) Throughput for 4,8 and 16 node respectively with varying Simulation Time

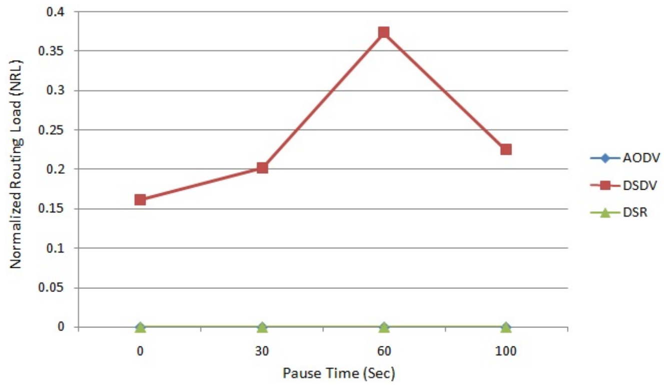
In Figure 3(a,b,c) it can be seen from the results, throughput of AODV is higher than DSR and DSDV since its routing overhead is less than others. The rate of packet received for AODV is better than the DSDV. The dropped packet for DSR is less than that of DSDV; AODV has no periodic updates exist in DSR. DSDV routing protocol consumes more bandwidth, because of the frequent broad casting of routing updates. While the AODV is better than DSDV as it doesn't maintain any routing tables at nodes which results in less overhead and more bandwidth.



4.a



4.b



4.c

Fig:4(a,b,c) Normalized routing load for 4,8 and 16 node respectively with varying Simulation Time

In Figure 4(a,b,c) it can be seen from the results, normalized routing load is minimum at DSR and also is AODV produce low results in compare to DSDV because the normalized routing load is defined as the fraction of all routing control packets sent by all nodes over the number of received data packets at the destination nodes.

B. Fixed Simulation Times but Varying Number of Nodes

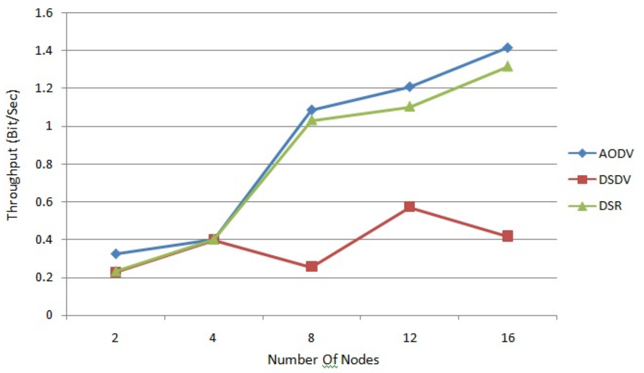


Fig.5. Variation of Throughput by varying number of nodes

In Figure 5 It can be seen that, the Throughput of AODV is higher than DSR and DSDV since its routing overhead is less than others. The rate of packet received for AODV is better than the DSDV. The dropped packet for DSR is less than that of DSDV; AODV has no periodic updates exist in DSR. DSDV routing protocol consumes more bandwidth, because of the frequent broad casting of routing updates. While the AODV is better than DSDV as it doesn't maintain any routing tables at nodes which results in less overhead and more bandwidth.

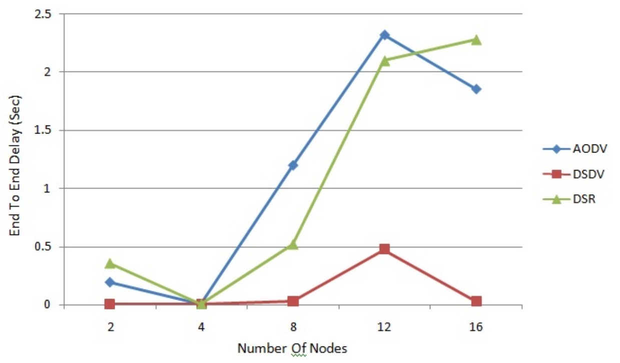


Fig:6. Variation of End to end delay by varying number of nodes

In Figure 6. It can be seen that, the End to End delay of AODV is higher than DSR and DSDV. In END to End Delay, DSDV which is a table driven proactive routing protocol completely wins over the on demand reactive routing protocols DSR and AODV .Since DSDV proactively keeps the routes to all destination in its table it does not have to initiate the route request process as frequently as in DSR and AODV while sending packets. Hence on average DSDV clearly has less delay. It can be observed that AODV is the worst protocol in terms of delay.

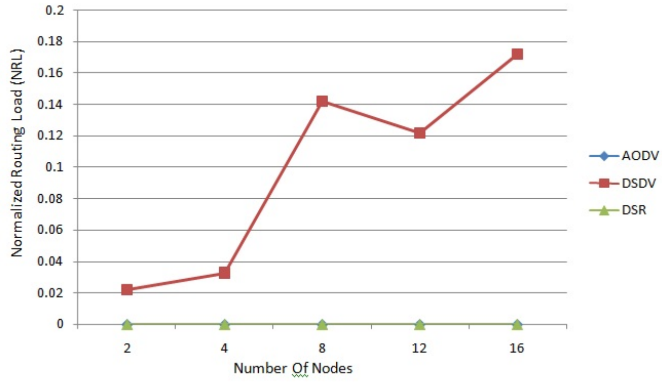


Fig.7. Variation of Normalized routing load by varying number of nodes

In Figure 7 it can be seen that, the Normalized Routing Load of DSDV is higher than DSR and AODV and the load in routing is minimum at DSR and AODV because the normalized routing load is defined as the fraction of all routing control packets sent by all nodes over the number of received data packets at the destination nodes. Routing load of DSDV is maximized so it is not fine.

V. CONCLUSION

Here the performance about three different commonly used mobile ad hoc routing protocols namely AODV, DSR and DSDV by increasing numbers of nodes and push time using NS-2 has been evaluated. The performance analysis depend on throughput, average end to end delay and normalized routing load. It can be concluded that DSR gives good performance when comparing to AODV, and DSDV in terms of packet delivery ratio and end to end delay. AODV gives better performance than DSR and DSDV in terms of throughput.

REFERENCES

[1] Sachin Kumar Gupta and R K Saket,"Performance Metric Comparison of AODV and DSDV Routing Protocols in MANETs Using NS-2",IJRRAS,Vol. 7,Issue No.3, June 2011.

[2] Manveen Singh Chadha and Rambir Joon Sandeep, "Simulation and Comparison of AODV, DSR and AOMDV Routing Protocols in MANETs", International Journals of Soft Computing and Engineering (IJSCE), Vol.2, July 2012.

[3] S. Vanthana and Dr. V.Sinthu Janita Prakash, "Comparative Study of Proactive and Reactive Adhoc Routing Protocols Using NS2", IEEE – World Congress On Computing and Communication Technologies, 2014.

[4] Dr. D. Sivakumar and B. Suseela, "A Survey of Routing Algorithms For MANET ", IEEE-International Conference On Advances in Engineering, Science, Management (ICAESM-2012).

[5] P. Manickam, T. Guru Baskar, M.Girija and Dr.D. Manimegalai, "Performance Comparison Of Routing Protocols In Mobile Ad Hoc Networks",IJWMN, 2011.

[6] C.E. Perkins and P. Bhagwat, “Highly Dynamic Destination Sequenced Distance-vector Routing (DSDV) for Mobile Computers”, ACM SIGCOMM ’94 Conference,1994.

[7] Thakur Anuja, Sharda Patel, and Ashok Verma. “Performance Evaluation of AODV, DSDV and DSR Routing Protocols using NS- 2 Simulator”, International Journal Of Engineering Science & Research technology, 2014.

[8] Kavita Pandey and Abhishek Swaroop, “A Comprehensive Performance Analysis of Proactive, Reactive and Hybrid MANETs Routing Protocols”, IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 6, No 3, November 2011.

[9] M Bouhoma,H Bentaouit , and A.Boudhir,”Performance comparison of Ad hoc Routing Protocols AODV and DSR”,IEEE,2009.

Comparison of Invasive Weed Optimization (IWO) and Particle Swarm Optimization (PSO) in improving power system stability by UPFC controller employing a Multi-objective approach

Mohammad Shoaib Shahriar[1*], Md Shafiullah[1], Mohammed Afzal Asif[2], Md Mahmudul Hasan[2], Ahmed Ishaque Imtinan Rajgir[2]

[1]Department of Electrical Engineering, King Fahd University of Petroleum & Minerals Dhahran, Saudi Arabia

[2]Department of Electrical and Electronic Engineering, Independent University, Bangladesh

*shoebeee05@gmail.com

***Abstract*— In the arena of power system, small signal stability is one of the most important issue to be settled down. In relatively weak tie line interconnections, these low frequency oscillations are observed. Use of Flexible AC Transmission System (FACTS) devices are playing most important roles in damping out of those oscillations. Unified Power Flow Controller (UPFC), one of the key members of FACTS family are being used in long and high voltage transmission networks of modern power system. Tuning of the UPFC parameters in real time power system application has been considered in making the optimization problem. Two of the novel optimization approaches namely Particle Swarm Optimization (PSO) and Invasive Weed Optimization (IWO) has been proposed in this work to optimize the parameters of UPFC and the efficacy of the optimizers in tuning UPFC has been investigated and compared through time domain analysis of different machine parameters such as variation of rotor angle and angular frequency with respect to time.**

Keywords*— Particle Swarm Optimization; Invasive Weed Optimization; Eigenvalue; FACTS; Multi-objective Optimization; Power System Stability; UPFC.

I. INTRODUCTION

Weakly connected tie lines are one of the common scenarios in interconnected large power systems which results low frequency oscillations. These low frequency and small magnitude oscillations sometimes last for a longer period of time [1]. Controlling the excitation is one of the most common ways to control power system stability. Automatic Voltage Regulator (AVR) is one of the widely used exciters. Power System Stabilizer (PSS) is another extensively used tool in damping out the low frequency oscillations in order to improve system stability [2-3] throughout the globe. However, using such conventional damping mechanism could not be able to provide sufficient damping for inter-area oscillations, sometimes.

After 1980s, use of the technology named flexible AC transmission systems (FACTS) [4-7] has been increased significantly because of the advancement of power electronic devices. Using such FACTS-based controller provides a strong solution in power system instability even for the case of inter-area modes. Among these FACTS devices, one of the 2nd generation item is Unified Power Flow Controller (UPFC). Beside of its primary function of controlling power flow, it could also be used for enhancement of transient stability, voltage drop support, minimizing power loss, damping oscillations out and so on [7-9]. Proper control strategies applied on UPFC converters are the main factors here for stability enhancement. The supplementary control of UPFC has a significant effect on the system when connected to the interconnected power systems. But lack of proper coordination, UPFC and PSSs may cause some sort of unwanted interactions between them which might cause the system destabilization later. So, the coordination among PSS and FACTS device controllers [10-11] has been researched a lot to improve the performance of the system in total. Extensive amount of studies has been made to explore different optimization techniques in finding out the best possible solution in recent years.

Particle Swarm Optimization (PSO) is a popular population based optimization technique that shares much resemblance to Genetic Algorithm (GA) but operates in a rather 'constructive matter' in resolving optimization problems [12-14]. And Invasive Weed Optimization (IWO) is a meta-heuristic optimization algorithm, first proposed in [16-18], which represents the ecological behavior of the colonizing weeds in solving different kinds of non-differentiable, complex, nonlinear numerical optimization problems. Both the techniques are widely used in different application fields for their stochastic search property. Problem-solving ability of PSO and IWO is not that much sensitive to the selection of initial value of the decision variable parameter.

In this paper the optimization capability of IWO and PSO is proposed and compared in designing the PSSs tuning parameters. The optimization problem has been formulated as a constrained optimization problem. Two Eigen value based objectives has been used to improve system damping. As system model, a single machine infinite bus (SMIB) power system connected with FACTS device UPFC has been taken to investigate the performance. Both the time domain simulation and Eigen value analysis has been performed.

II. PARTICLE SWARM OPTIMIZATION (PSO) ALGORITHM

Particle swarm optimization (PSO) is an optimization method with its algorithm based on swarm intelligence. The concept was laid out at first by Doctor Kennedy and Eberhart in 1995 [14]. It is applied on a population of particles where each particle has a random value for velocity and position. A desired value shall be given and the particles will rearrange themselves by updating the contained value each particle has. This process will go through several iterations until any particle's initialized value gets closest to the desired value.

PSO has characteristics for which it is the first choice for many applications. PSO is easy to apply and its computational expense is low since its memory occupation and processing speed requirement is low. The working steps are described below [13, 15]:

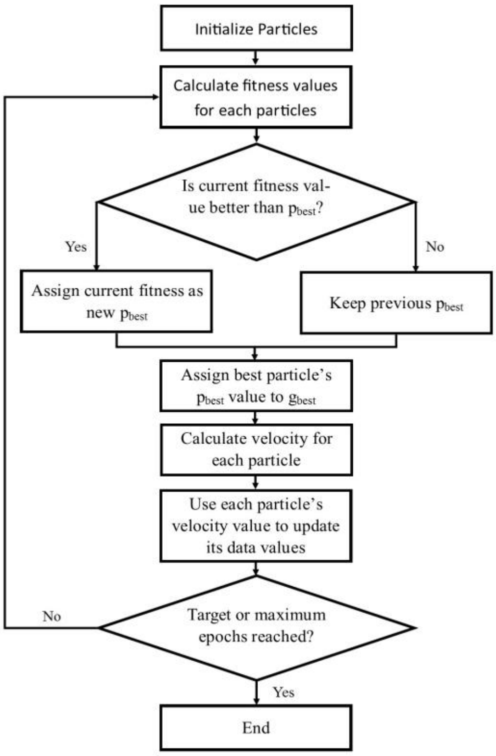


Fig 1 A flowchart representation of PSO Algorithm

A. ***Initialization***

All particles' velocity(v) and position(p) are set randomly.

B. ***Velocity Updating***

The velocity of each particle is updated according to the formula stated below [12]:

$$\vec{v}_n = w\vec{v_n} + c_1R_1(\vec{p}_{best} - \vec{p}_n) + c_2R_2(\vec{g}_{best} - \vec{p}_n) \quad (1)$$

p_n and v_n are position and velocity of a particle n. p_{best} is the best individual value of position of the particle n found so far and g_{best} is the best position value found in the entire population. w is a constant term that controls the flying dynamics. R_1 and R_2 have values ranging from 0 to 1. c_1 and c_2 are controlling factors. v_n should be checked if it is within the range.

C. ***Position Updating***

The position of particles is updated according to the formula stated below [12]:

$$\vec{p}_1 = \vec{p}_n + \vec{v}_n \quad (2)$$

After every update p_n should checked if it is within the range.

D. ***Memory Updating***

p_{best} and g_{best} are updated when the condition is met [12].

$$\vec{p}_{best} = \vec{p}_n \qquad if\ f(\vec{p}_n) > f(\vec{p}_{best}) \quad (3)$$

$$\vec{g}_{best} = \vec{g}_n \qquad if\ f(\vec{g}_n) > f(\vec{g}_{best}) \quad (4)$$

f(x) is the function used to find the best value.

E. ***Termination Checking***

Steps 2 and 4 are repeated until the terminating conditions are met. When terminating conditions are met, the algorithm returns the value of g_{best} and $f(g_{best})$.

III. INVASIVE WEED OPTIMIZATION (IWO)

Invasive Weed Optimization (IWO) is an algorithm based upon the practical behavior of a colony of weeds [17, 18]. It imitates the random spatial distribution of the seeds and selective elimination method found in the natural colony of weeds. Invasive weed optimization algorithm progresses in four steps:

A. ***Initialization*****:**

A fixed set of solutions or weeds are initialized and addressed at equal separations of half λ. The initial number of weeds establishing a colony can vary with the nature of the problem.

B. ***Reproduction*****:**

This step is based on the generation of seeds. Every weed in the colony produces a count of seeds depending upon its ability. Thus, the weakest weed gives the least number of seeds (S_{min}) and the strongest weed has the highest seed count (S_{max}).

C. ***Seed Dispersal*****:**

The basis of the algorithm is set upon the arbitrary distribution of the seeds within the colony. The dispersed seeds are laid over the *d*-dimensional search space randomly with zero mean. The extent of standard deviation from the mean value can be acquired from the following equation:

$$\sigma_{iter} = \frac{(iter_{\max} - iter)^n}{(iter_{\max})^n}(\sigma_{initial} - \sigma_{final}) + \sigma_{final}$$

According to the equation, the standard deviation decreases from $\sigma_{initial}$ (highest standard deviation) to σ_{final} (lowest standard deviation). Hence, the spreading of seeds is such that, a stronger weed will disperse higher number of seeds around it whereas a weaker mother-plant will have relatively less seeds surrounding it.

D. *Selective Exclusion:*

Abiding the standard deviation equation of seed dispersal, after a few iterations the number of seeds generated by fitter weeds will be greater compared to that of the undesired less fit weeds. With each iteration, as the seeds grow into plants the rate of weed generation descends. This is because the population of plants reaches a maximum limit, pop_{max}. This is closely followed by the elimination process. In this process, the number of weaker plants get reduced in competition with the fitter offspring of the stronger mother weeds, to the point annihilation.

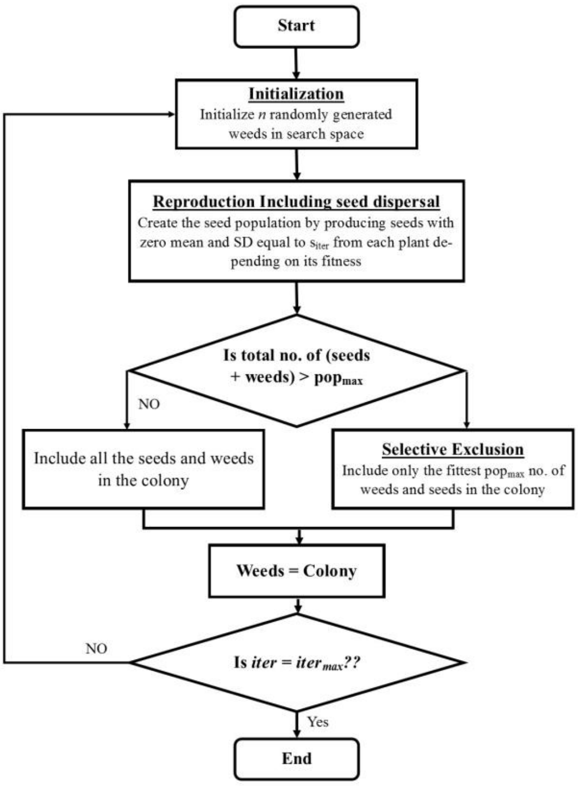


Fig 2 A flowchart representation of IWO Algorithm

IV. POWER SYSTEM DYNAMIC MODEL

A. *SMIB System Model*

The power system model [17] that has been chosen here to analyze has got a synchronous generator. Through a transmission line, this source end is connected to an infinite bus via two transformers (excitation transformer and boosting transformer). Two three-phase voltage source converters are attached with a capacitor (DC link). m_E, m_B, δ_E and δ_B are the four control signals of UPFC. Described SMIB system equipped with an UPFC is shown in Fig. 1.

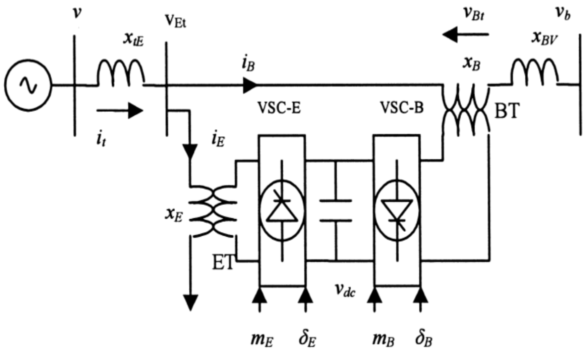


Fig 3: SMIB system with UPFC

Dynamic system model has been used here for small signal stability in this work. Parameters for system initialization are shown in Appendix. In order to work with stability and control of power system oscillations, linearization will be done later and the linearized model will be used.

The SMIB system nonlinear model shown in Fig 3 can be represented by the following three differential equations (5-7) [15]:

$$\dot{\delta} = \omega_{base}(\omega - 1) \quad (5)$$

$$\dot{\omega} = \frac{1}{2H}[P_m - P_e - D(\omega - 1)] \quad (6)$$

$$\dot{E}_q^{'} = \frac{1}{T_{do}^{'}}[E_{fd} - E_q^{'} - (x_d - x_d^{'})i_d] \quad (7)$$

Where M, D, P_m and P_e are the inertia constant and damping coefficient, input and output power applied to the system, respectively. ω_{base} represents the synchronous speed. δ is rotor angle and ω represents speed. $E_q^{'}$ and E_{fd} are the generator internal and field voltages, respectively. x_d and $x_d^{'}$ are the d axis transient and sub transient reactance used in the system model.

$$P_e = v_d i_d + v_q i_q, v_t = \sqrt{\left(v_d^2 + v_q^2\right)}$$
$$v_d = x_q i_q = x_q\left(i_{iq} + i_{lq}\right)$$

B. *Exciter and PSS*

The excitation system model (IEEE ST1 type) is shown in Fig. 4. Automatic Voltage Regulator (AVR) is used here to give the excitation to the system. The Power System Stabilizer (PSS) contains a washout circuit as well as a lead-lag block along with a limiter. This excitation part can be represented by:

$$\Delta\dot{E}_{fd} = K_A((V_{ref} - v + U_{pss}) - E_{fd})\frac{1}{T_A} \quad (8)$$

Where T_A and K_A are the time constant and gain of the excitation system, respectively; V_T is the terminal voltage and V_{ref} is reference voltage. U_{PSS} is the control signal comes from PSS.

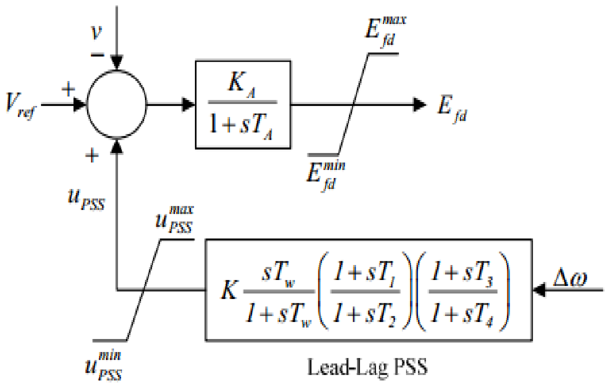


Fig. 4: Block diagram of the excitation system

C. The Dynamic Model of the UPFC

The UPFC model can be presented by the following differential equations [18] after applying Park's transformation in the system model:

$$\dot{v}_{dc} = \frac{3m_E}{4C_{dc}}(\cos\delta_E i_{Ed} + \sin\delta_E i_{Eq}) + \frac{3m_B}{4C_{dc}}(\cos\delta_B i_{Bd} + \sin\delta_B i_{Bq}) \qquad (9)$$

Where, C_{dc} is the DC link capacitance and V_{dc} is the voltage across that. The relations of the line currents with different parameters of excitation and boosting transformers are as follows:

$$i_{Ed} = \frac{x_{BB}}{x_{d\Sigma}}E_q' - \frac{m_E Sin\delta_E v_{dc} x_{Bd}}{2x_{d\Sigma}} + \frac{x_{dE}}{x_{d\Sigma}}(v_b Cos\delta + \frac{m_B Sin\delta_B v_{dc}}{2})$$

$$i_{Eq} = \frac{m_E Cos\delta_E v_{dc} x_{Bq}}{2x_{q\Sigma}} - \frac{x_{qE}}{x_{q\Sigma}}(v_b Sin\delta + \frac{m_B Sin\delta_B v_{dc}}{2})$$

$$i_{Bd} = \frac{x_E}{x_{d\Sigma}}E_q' + \frac{m_E Sin\delta_E v_{dc} x_{dE}}{2x_{d\Sigma}} - \frac{x_{dt}}{x_{d\Sigma}}(v_b Cos\delta + \frac{m_B Sin\delta_B v_{dc}}{2})$$

$$i_{Bq} = -\frac{m_E Cos\delta_E v_{dc} x_{qE}}{2x_{q\Sigma}} + \frac{x_{qt}}{x_{q\Sigma}}(v_b Sin\delta + \frac{m_B Cos\delta_B v_{dc}}{2})$$

D. UPFC based Damping Controller

Four control parameters (δ_B, m_B, m_E and δ_E) of UPFC can be used to control the damping torque. The structure damping controller based on UPFC is shown in Fig. 5, where control signal input (u) comes from UPFC. This controller used here could be considered as a lead/lag compensator.

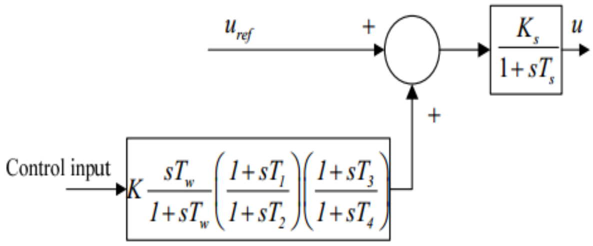


Fig. 5: UPFC with lead lag controller

E. Linearized Model

The linear model representation is needed for designing controller and analyzing system properly. By combining the equations 5-9 and then linearizing, we got the state space model of the system which can be presented as:

$$\dot{X} = A\Delta X + B\Delta U \qquad (10)$$

Where state vector $\Delta X = [\Delta\delta, \Delta\omega, \Delta E'_q, \Delta E_{fd}, \Delta V_{dc}]^T$,
Control vector $\Delta U = [\Delta U_{pss}, \Delta m_E\ \Delta\delta_E\ \Delta m_B\ \Delta\delta_B]^T$

$$A = \begin{bmatrix} 0 & \omega_b & 0 & 0 & 0 \\ -\frac{K_1}{M} & -\frac{D}{M} & -\frac{K_2}{M} & 0 & -\frac{K_{pd}}{M} \\ -\frac{K_4}{T'_{d0}} & 0 & -\frac{K_3}{T'_{d0}} & \frac{1}{T'_{d0}} & -\frac{K_{qd}}{T'_{d0}} \\ -\frac{K_A K_5}{T_A} & 0 & -\frac{K_A K_6}{T_A} & -\frac{1}{T_A} & -\frac{K_A K_{vd}}{T_A} \\ K_7 & 0 & K_8 & 0 & -K_9 \end{bmatrix}$$

$$B = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & -\frac{K_{pe}}{M} & -\frac{K_{pde}}{M} & -\frac{K_{pb}}{M} & -\frac{K_{pdb}}{M} \\ 0 & -\frac{K_{qe}}{T'_{d0}} & -\frac{K_{qde}}{T'_{d0}} & \frac{K_{qb}}{T'_{d0}} & -\frac{K_{qdb}}{T'_{d0}} \\ \frac{K_A}{T_A} & -\frac{K_A K_{ve}}{T_A} & -\frac{K_A K_{vde}}{T_A} & -\frac{K_A K_{vb}}{T_A} & -\frac{K_A K_{vdb}}{T_A} \\ 0 & K_{ce} & K_{cde} & K_{cb} & K_{cdb} \end{bmatrix}$$

V. OPTIMIZATION PROBLEM

In formulating the optimization problem, a multi-objective function based on Eigen value analysis has been chosen which involves the combination of two major decision taking parameters: damping factor and damping ratio. Main objective function J combines two separate objective functions J1 and J2 as follows [19]:

$$J = J_1 + \alpha J_2 \qquad (11)$$

$$J = \sum_{\sigma_{ij} \geq \sigma_0} (\sigma_0 - \sigma_{ij})^2 + a \sum_{\zeta_i \geq \zeta_0} (\zeta_0 - \zeta_i)^2 \qquad (12)$$

Where i is the index of Eigenvalues in the system and j is the index of system operating conditions. Real part of Eigenvalue σ_i determines the relative stability of the system in terms of damping factor margin. ζ_i is the damping ratio of the i^{th} Eigenvalue. The value of 'a' is the weight to combine J_1 and J_2. Value of 'a' has been chosen 10 for this work.

Optimization of the combined objective J, not only places the closed loop system poles in the left half plane of the jω axis but also reduces system oscillation. Thus, enhancement in the relative stability is achieved in the system.

Fig. 6 reflects the graphical representation of the selected objective functions in the field of system stability. If the objective function J1 has been taken only, the closed loop Eigen values will be forced to be placed in the indicated dashed region to the left of a line as shown in Fig. 6a. Similarly, choosing only J_2 as objective function will limit the maximum overshoot of the Eigen values within a particular dashed region as shown in Fig. 6b. When optimized with combined objective function J, the Eigen values generated by the system will be restricted within a particular D-shaped area as shown in the shaded portion of Fig. 6c.

So, the optimization problem can be stated as:

$$\textit{\textbf{Minimize}} \quad \textit{\textbf{J}}$$

Subject to

$K^{min} \le K \le K^{max}, \quad T_1^{min} \le T_1 \le T_1^{max}$

$T_2^{min} \le T_2 \le T_2^{max}, \quad T_3^{min} \le T_3 \le T_3^{max}$

$T_4^{min} \le T_4 \le T_4^{max}$

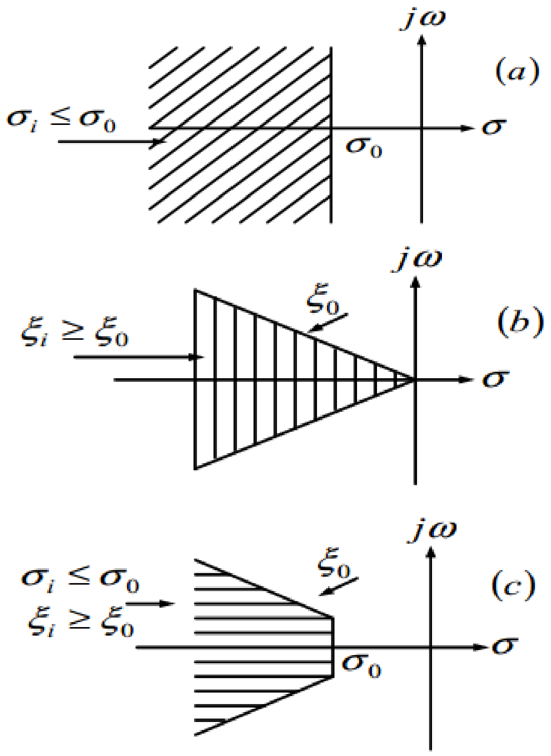


Fig 6: Eigen value locations for corresponding objective function

VI. SIMULATION RESULTS

One of the ways to check the stability of the system is to analyze the Eigen values of the system matrix. Looking at the symbol of the value of the real part of it, we can make a guess about the behavior of the whole system. We performed the Eigen value analysis for all three cases: with only conventional PSS tuning, using PSO and using IWO.

Table I: Eigenvalues Comparison for three different cases with Nominal Loading

Conventional PSS	With PSO	With IWO
-9999.6839+0i	-9998.6308+0i	-9996.7984+0i
-93.274993+0i	-93.592009+0i	-94.19553+0i
-6.2503608+0i	-2.1868007+6.09692i	-5.9619+8.40415i
0.13450675+4.681183i	-2.1868007-6.09692i	-5.9618772-8.40415i
0.13450675-4.681183i	-3.6749997+0i	-1.78201+1.19083i
-1.5930012+0i	-0.24652855+0i	-1.78201-1.19083i
-0.61662513+0.1326385i	-1.5035978+0i	-1.4802187+0i
-0.61662513-0.1326385i	-1.5006631+0.05676i	-0.24190767+0i
-0.26664725+0i	-1.5006631-0.056766i	-100+0i

Table I clearly shows the use of PSO and IWO in finding out the optimal set of UPFC parameters. Random selection of constant parameters results some positive real part Eigen values which actually represents the unstable behavior of the system. But all the Eigenvalues, after applying PSO or IWO are with negative real parts that satisfies the stable condition of the system. In comparing between the two, IWO performs much better than PSO as it provides most of the values more negative than the corresponding values of PSO.

The time domain simulation results for six seconds after being subjected to physical disturbances are shown in the following figures. Fig. 7 shows the DC link voltage characteristic which clearly depicts, UPFC tuned by PSO and IWO shows stable behavior whereas with the conventional one leads towards instability.

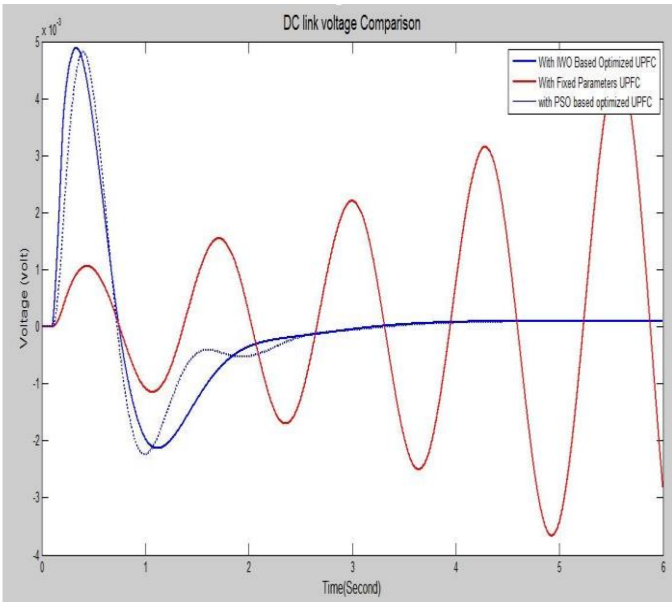


Fig 7: V_{dc} vs time curve with and without PSO and IWO optimization

Figs. 8 and 9 show the variations of rotor angle and angular speed of the machine with respect time. Both figures show PSO and IWO optimize UPFC stabilize the system whereas conventional one leads towards unstable behavior. Maximum value of oscillation and time needed to be optimized defer for both the optimization techniques but both of the techniques showed the robustness and accuracy in tuning the PSS parameters.

The total simulation was done by MATLAB simulation software. Table II represents a summarized view of comparison between PSO and IWO while performing this test. The value of objective function which is to be minimized is 15.78 with IWO which is better than the value obtained with PSO optimization. PSO is much quicker than IWO in working speed while IWO needs less number of iterations to reach the optimal solution. Optimized PSS parameters are also presented which are essential.

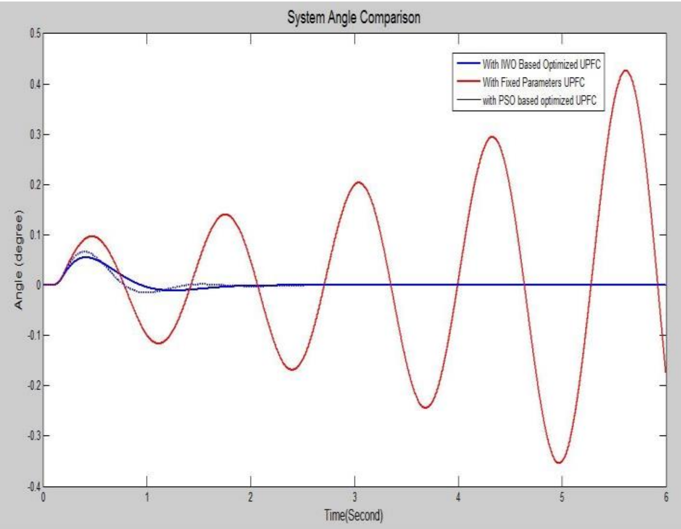


Fig 8: δ vs time curve with and without PSO and IWO optimization

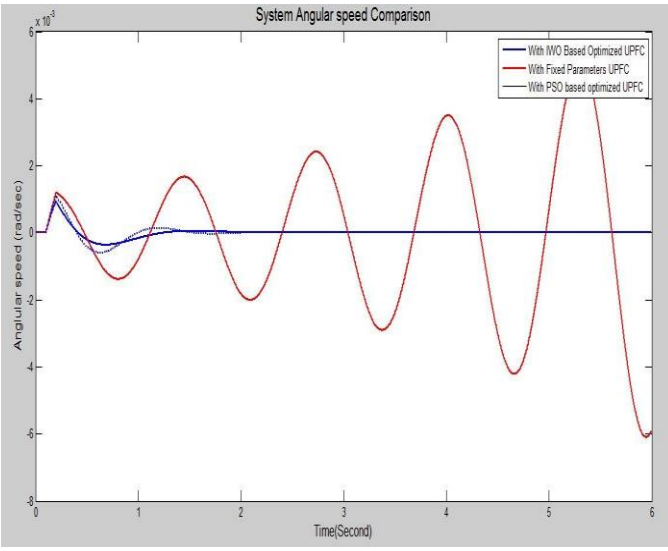


Fig 9: ω vs time curve with and without PSO and IWO optimization

The overall performance of two different robust optimizers are summarized in the table II below:

Table II: Comparison between PSO and IWO

	IWO	PSO
Objective function value	15.7840	19.7025
Time needed	762.4335 seconds	11.19236 seconds
No of iteration	35	100
Optimized parameters	Kc=19.6064 , T1=0.2520 , T2=0.3421, T3=0.3954, T4=0.3210	Kc=15.7119, T1=0.2823, T2=0.2873, T3=0. 0.5639, T4=0.5623

VII. CONCLUSION

In order to deal with the stability issue of power system, designing the controller demands a lot of attention. The optimization problem in this work which is solved by two of the new techniques of optimization named PSO and IWO with the multi-objective function based on Eigen value analysis. The performance of the proposed UPFC controller is demonstrated subjected to a particular kind of disturbance. The Eigen value analysis as well as the nonlinear time domain simulation presented in result section show the effectiveness of the proposed controllers which uses the multi-objective function. It also reflects the ability to provide good amount of damping in the arena of low frequency oscillations. In choosing one of the techniques from two, it's difficult to decide as both the optimization techniques show remarkable feats to the conventional PSS, with IWO providing a better set of solutions and PSO proving to be very fast in operation.

APPENDIX

The parameters of the system used in optimization results of the UPFC are:

Generator and transmission line parameters: D=0, M = 8 MJ/MVA, T'_{d0} = 5.044 s, x_d= 1.0 pu, x'_d = 0.3 pu, x_q= 0.6 pu, ω_b =377 rad/s and X_L = 0.1 pu

Excitation system and transformer parameters: T_A=0.01s, K_A=100, X_E= 0.1 pu, X_T= 0.1 pu and X_B = 0.1 pu

UPFC parameters: m_E= 0.7667, m_B = 0.96, δ_E =68.113 deg and δ_B =41.118 deg

DC link parameters: C_{DC} = 1.2 pu and v_{DC}= 2 pu

Operating condition: P_e = 0.98 pu, Q_e= - 0.1670pu, V_b = 1 pu and V_t = 1.05 pu

REFERENCES

[1] P. Kundur, Power System Stability and Control, McGraw-Hill, 1994.

[2] E. Larsen, D. Swann, “Applying Power System Stabilizers Part I: General Concepts”, IEEE Transactions on Power Apparatus and Systems, Vol. 100, No. 6, pp. 3017-3046, 1981.

[3] M. Eslami, H. Shareef and A. Mohamed, “Application of artificial intelligent techniques in PSS design: a survey of the state-ofthe-art methods, Prz. Elektrotechniczny , Vol. 87, No. 4, pp. 188-197, 2011.

[4] M. Eslami, H. Shareef and A. Mohamed, “Application of PSS and FACTS Devices for Intensification of Power System Stability, Int. Rev. Electr. Eng., Vol. 5, No. 2, pp. 552-570, 2010.

[5] M. Eslami, H. Shareef and A. Mohamed, “A Survey on Flexible AC Transmission Systems (FACTS), Prz. Elektrotechniczny, Vol. 88, No. 1A, pp. 1-11, 2012.

[6] J. Shea, "Understanding FACTS-concepts and technology of flexible AC transmission systems," Electrical Insulation Magazine, IEEE, Vol. 18, No. 1, pp.46, 2002.

[7] Hingorani, N. G. "Flexible AC Transmission", IEEE Spectrum, Vol. 30, pp. 40-45, April 1993.

[8] Wartana, I.M. and Agustini, N.P. "Optimal Placement of UPFC for maximizing system loadability and minimizing active power losses in system stability MARGINS by NSGA-II", international conference on electrical engineering and informatics (ICEEI), pp. 1-6, july 2011.

[9] H. Wang, “Applications of modelling UPFC into multi-machine power systems, IEE Proceedings-Generation, Transmission and Distribution, Vol. 146, No. 3, pp. 306-312, 1999.

[10] M. Eslami, H. Shareef and A. Mohamed, “Coordinated design of PSS and TCSC controller for power system stability improvement. In The 9th International Power and Energy Conference (IPEC), Singapore, pp. 433-438, 2010

[11] M. Eslami, H. Shareef and A. Mohamed, “Optimization and Coordination of Damping Controls for Optimal Oscillations Damping in Multi-Machine Power System", Int. Rev. Electr. Eng., Vol. 6, No. 4, pp. 1984-1993, 2011.

[12] Dr. Karl O. Jones, "Comparison Of Genetic Algorithm And Particle Swarm Optimisation", Internation Conference on Computer Sciences and Technologies (CompSysTech), pp. 1-6, 2015.

[13] Ashik Ahmed, B M Ruhul Amin, “Performance Comparison of Invasive Weed Optimization and Particle Swarm Optimization Algorithm for the tuning of Power System Stabilizer in Multi-machine Power System”, International Journal of Computer Applications (0975 – 8887) Volume 41– No.16, March 2012.

[14] Dian P. Rani, Siti M. Shamsuddin, Siti S. Yuhaniz, “Particle Swarm Optimization: Technique, System and Challenges”, International Journal of Computer Applications (0975 – 8887) Volume 14– No.1, 2012.

[15] H. Madivada, C.S.P. Rao, “An Invasive Weed Optimization (IWO) approach for Multi-Objective Job Shop Scheduling Problems (JSSPs)”, Internation Journal of Mechanical Engineering and Technology (0976 - 6359) Volume 3, pp. 627-637, December 2012.

[16] K. Deb, A. Pratap, S. Agarwal, “A fast and elitist multiobjective genetic algorithm: Nsga-II, IEEE Trans. Evol. Comput. Vol. 6, pp. 182–197, 2012.

[17] Mehrabian, A.R. and Lucas, C. (2006) ‘A novel numerical optimization algorithm inspired from weed colonization’, Ecological lnformatics, Vol. 1, No. 4, pp.355–366, 2006.

[18] Sepehri Rad, H. and Lucas, C. (2007) ‘A recommender system based on invasive weed optimization algorithm’, in Proc. IEEE Congress on Evolutionary Computation, pp.4297– 4304, 2007.

[19] Al-Awami A. T. - Abdel-Magid Y. L. - Abido M. A. : “Simultaneous Stabilization of Power System Using UPFC-Based Controllers”, Electric Power Components and Systems, Vol. 34, pp. 941–959, 2006.

[20] H. Shayeghi, S. Jalilizadeh, H. Shayanfar, A. Safari, “Simultaneous coordinated designing of UPFC and PSS output feedback controllers using PSO”, Journal of Electrical Engineering, Vol. 60, No. 4, pp. 177–184, 2009.

[21] A. Ajami and M. Armaghan, " Application of Multi-Objective PSO Algorithm for Power System Stability Enhancement by Means of SSSC", International Journal of Computer and ElectricalEngineering, Vol. 2, No. 5, pp. 1793-8163, 2010.

Context Aware Energy Allocation by Auction Based Method in Wireless Sensor Networks

Imam Ush Shaheed, Sadia Sabrin Nodi and Muhidul Islam Khan
Department of Computer Science and Engineering
BRAC University
Dhaka, Bangladesh
Email: iushaheed@gmail.com, sadiasabrin100@gmail.com and muhit.islam.khan@gmail.com

***Abstract*—Wireless sensor networks refer to hundreds and even thousands of small tiny devices called sensor nodes distributed autonomously to observe physical or environmental parameters like temperature, pressure, vibration and motion at different locations such as landslides. Every node in a sensor network usually equipped with one sensor, a wireless communications device like radio transceiver, a small microcontroller, and an energy supply, a battery. Since the nodes are battery operated energy plays a vital role. The application of the WSN involves several fields, like military battleground, fire detection, and other extreme environments. In these situations, it is troublesome to replace the dead nodes caused by energy depletion with new ones to provide energy for the system. Therefore, making sensor nodes operating as long as possible is the main method to maximize the lifespan of the sensor network. Context aware task allocation/energy allocation is an important issue for maximizing the lifetime of the network. In this research our goal is to minimize the wastage of energy and to maximize the usage by context aware energy allocation. We develop a context aware energy allocation algorithm based on First-Price auction method. Our simulation results show that our proposed method provides better results in terms of energy consumption comparing with the other existing methods.

Keywords—Wireless Sensor Networks; Context Aware Energy Allocation; Auction based method; First-Price auction.

I. INTRODUCTION

A wireless sensor network is a group of specialized transducers with a communications infrastructure for monitoring and recording conditions at diverse locations. Commonly monitored parameters are temperature, humidity, pressure, wind direction and speed, illumination intensity, vibration intensity, sound intensity, power-line voltage, chemical concentrations, pollutant levels and vital body functions [12]. WSN usually is battery powered which is very energy consuming. At present efficiently allocation of energy is one of the main challenges of WSNs [5]. To use the energy efficiently researchers have been working on WSN power/energy allocation. They are emphasizing on sensing, communication, computation and energy harvesting. In general, the communication consumes huge number of energy.

In our research we are working to find out an efficient algorithm/method of energy allocation which will be context aware and power efficient for WSN.

Context Aware means to behave dynamically with the environment and to act in a way that the situation demands. Suppose for a sensor node, it needs to be in sleep mode if there is nothing to detect or if a node needs to track something it will be activated immediately.

Generally, a wireless sensor network consists of three main components: Nodes, Gateways and Software [13]. The Nodes consist of several sensors are used to monitor assets or environment we are working on. The acquired data are transmitted to the gateway, which is connected to a host system where data collection, processing, analyzing and measurement data are performed using a software. Routers are a type of node that is used to expand WSN distance and its area of implementation.

There are many different kind of tiny sensors which is used in various purposes like in military issues, shopping malls, building utilities, border areas, forests security etc. which creates a smart environment from where we can collect several data. These sensors work in different phases according to the need of the environment. A single sensor might do multiple works such as sensing, broadcasting, transmitting, bidding etc. The main challenge in this process is to keep those sensors alive for the maximum period of time and get maximum output [1].

We propose an auction based method for context aware energy allocation in a WSN. We consider a clustered based WSN for various task allocations to the sensor nodes in a way that the energy is optimized and the performance is maximized. We consider a set of tasks/jobs. Based on a utility function and using First-Price auction cluster head of each cluster helps to allocate the jobs to the sensor nodes so that the network lifetime maximizes.

Rest of the paper is organized as follows. Section II describes the related works. Problem formulation is mentioned in Section III. Section IV has the description of proposed method. Section V has simulation-results and section VI concludes the paper with the future direction of our work.

II. RELATED WORK

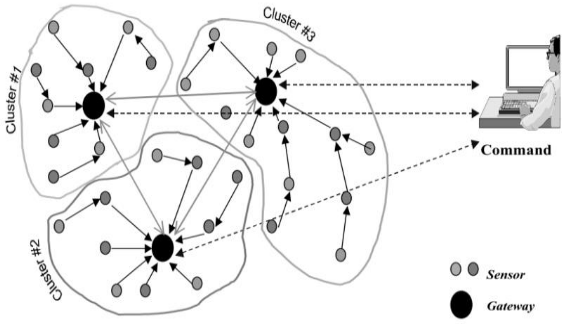


Fig. 1. Multi-gateway clustered sensor network [8]

A. Sinha et al. [4] proposed a prediction based method for task allocation. They have used power aware sensor node model which works on increasing latency and decreasing power consumption. The sleep states are differentiated by power consumption, the overhead required is going to sleep and the wake-up time. In general, a deeper sleep state consumes less power and has a longer wakeup time. An event occurs when a sensor node picks up a signal with power above a predetermined threshold. Here the calculated threshold level depends on the basis of average event rate. So the sleep states/wake up of the nodes depends on the probability of the occurrence of event. As the sleep time and wakeup time is based on the prediction of previous event, in worst case 2 things can be happened.

1) Events may occur during sleep time therefore they might have missed events.
2) No event may occur during wake time therefore the whole energy is being wasted during the period of time.

Compare to this prediction based model, our algorithm runs in such a way where there is a less possibility of wastage of energy and also less missed events as a node is being awaken only if the event occurs. This makes our work unique comparing with the existing approaches.

III. PROBLEM FORMULATION

Some applications of sensor networks include tracking the movements of birds, small animals, environmental monitoring in marine, soil, and atmospheric contexts, forest fire detection, meteorological or geophysical research, pollution study etc. There are some places in the world like forests and border area where it is dangerous to people to go and collect information's but which is needed. In these cases, a WSN plays a vital role. Since the nodes are battery operated, allocating energy is an important issue. The application of the WSN involves several fields, like military battleground, fire detection, and other extreme environments. Considering these situations, it is difficult to replace the dead nodes or their batteries. Therefore, making sensor nodes operating as long as possible is the main method to maximize the lifespan of the sensor network. Context aware energy allocation is an important issue for maximizing the lifetime of the network. In our research we focused on minimizing the wastage of energy and by this maximizing the usage by context aware energy allocation.

A single sensor node does many things such as capturing temperature, humidity, vehicular movement, lightning condition, pressure, soil makeup, noise levels etc. In our paper we consider our sensors can perform four tasks.

1) Receive
2) Transmit
3) Sense
4) Bidding

Our goal is to allocate these tasks to the sensor nodes in a way that the network lifetime increases.

IV. PROPOSED METHOD

To distribute the jobs of the sensors we divide the whole process in to three sub processes:

1) Cluster Head (CH) selection algorithm
2) Job allocation algorithm
3) Algorithm for selected node

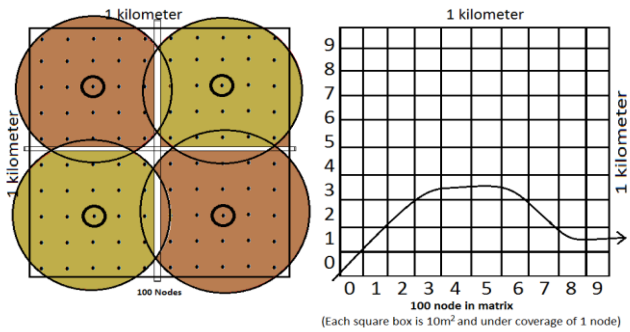


Fig. 2. Sensor node representation in 10x10 matrix

If we take an area of 1000 miter consist of 100 sensor nodes in 10/10 matrix. we take 4 main sensor nodes called main sensor nodes (these are not any of cluster heads. These are just 4 main sensor nodes). These four sensor nodes will always be in ON mode. Other nodes will be in sleep mode. The main sensor nodes will have the information of other nodes (such as location). These nodes always sense if there any substance is present in their range or not (if we implement this system in a border area to monitor illegal entry, substance can me defined as human from other side of the border, or if we implement this system to count tigers in a forest the tigers are the substances). We have assumed, the battery power of these main 4 sensor nodes are higher than the other nodes so that these nodes will be alive till the end.

Cluster Head (CH) selection:

When one of the four main sensor nodes detects a substance in its range it calculates the area and make the nearest node of the substance Cluster Head(CH) and follows to the second algorithm (Job allocation algorithm).

Algorithm 1 ClusterHeadSelection algorithm.

Input: ***matrixOfAllNode***, ***fourMainSensorNodes***
Output: ***Cluster_head***, ***AdjacentNodes***
Step 1: Start
Step 2: ***FourMainSensorNodes*** in 4 edges start sensing
Step 3: ***if FourMainSensorNodes[x]*** detects anything
i. Detect nearest node from that location
ii. Make it ***cluster_head***
iii. Initiate ***jobAllocationAlgorithm***
iv. go to **Step 3**
Step 4: exit

Job allocation algorithm:

In this part the cluster head selects its adjacent nodes and sends them a signal to initiate bidding. The selected nodes then bid to the cluster head according to their battery life and distance based on this bidding function.

Cluster heads gets all the bidding of all the adjacent nodes. It then finds out the two highest bidders. The highest bidder gets the job to sense and transmit and the second highest bidder remains in the queue as backup and send all the other bidders a message to terminate or sleep. If the highest bidder fails to complete its job, then the backup node continues the rest of the job. If the currently active node sends a message to terminate signal, it goes back to the cluster head selection algorithm if not it goes to the third algorithm (Algorithm for selected node).

Algorithm 2 Job allocation algorithm among bidders.

---

i. Phase I: INITIATE BIDDING
a. Select bidders (Neighbors)
b. Send "initiate bidding" signal to bidders

ii. Phase II: SETTING PRIORITY USING BIDDING FUNCTION
i. $B_{i\ [\ i=1,2,3,4,5,6,7,8]}$
ii. Using bidding function

$$\boxed{\frac{\mathbf{1}}{\boldsymbol{distance}(\boldsymbol{head},\boldsymbol{adj}[\boldsymbol{x}][\boldsymbol{y}])}+\boldsymbol{b\_Level}[\boldsymbol{x}][\boldsymbol{y}]}$$

iii. Saves returned array: ***priorityArray[][]***, stored high to low priority

iii. Phase III: WORK ALLOCATION
a. Head knows the winner/highest bidder
b. Assign the highest bidder to the job and 2nd highest as backup
priorityArray[0] → got the job
priorityArray[1] → in power saver as backup
c. Send all other bidders sleep/terminate signal

iv. **Phase IV: RECOVERY MODE**
a. ***if*** current node, ***priorityArray[0],*** fails initiate backup, ***priorityArray[1]***
b. ***priorityArray[1] → got the job***
c. recover **priorityArray[0]** then → in power saver mode as backup

v. Phase V: TERMINATION
a. ***if*** currently active node sends "**terminate signal"**
i. Send terminate signal to active and backup node
ii. Terminate (go back to ***ClusterHeadSelection***)

Algorithm for selected node:

The highest bidder will then check if the substance is in range and if DTT<0 or not. (DTT: Data to Transmit). If it is true, Then the bidder will start sensing and storing data.

If DTT >=1 the bidder will then transmit all the stored data to gateway and initiate the DTT in to 0.

If it doesn't match with any of the conditions, it means the substance is no more in the range it will transmit all the data to gateway and initiate the DTT into 0 again. Lastly it will send a signal to the cluster head to terminate its job. Thus the sensors will work according to the algorithm if it senses any substance into its range.

Algorithm 3 Algorithm for selected node. And backup node

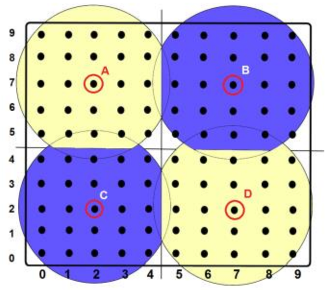
//DTT→ Data to Transmit
Step 1: Start
Step 2: check ***if*** the object ***isInRange ()*** and ***DTT***<1
Step 3: if true continue sensing and storing data of the object
Step 4: else
if **DTT** >=1
i. transmits stored data to ***gateway***
ii. ***DTT*** → 0;
iii. Go to Step 2
else // object is not in range now
i. transmits stored data to ***gateway***

ii. ***DTT*** → 0;

iii. Send terminate signal to ***cluster_head***

iv. Sleep.

Step 5: exit

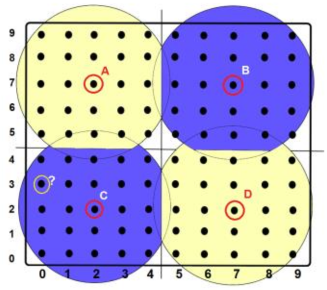


STEP 1

Node A B C D remains active

They will initiate **ClusterHeadSelection Algorithm** if detects any object in their area

#Suppose there is an object detected by C node in C region



STEP 2

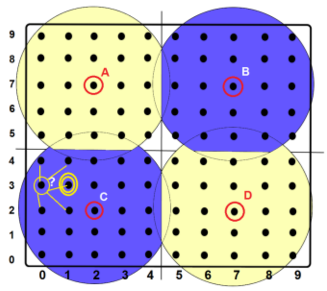
Algorithm 1:
Cluster Head Selection Algorithm

C node detects '?' object

Select **node(0,3)** as nearest node of object

make **node(0,3)** cluster_head

using this cluster_head **jobAllocationAlgorithm** starts



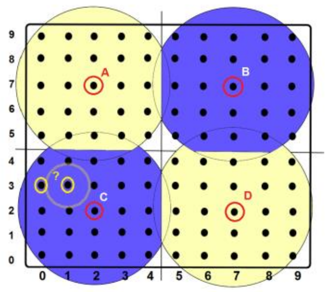
STEP 3

Algorithm 2 :
Job allocation algorithm

Initiate bidding among neighbors

$$\frac{1}{distance(head, adjecent[x][y])} + batteryLevel[x][y]$$

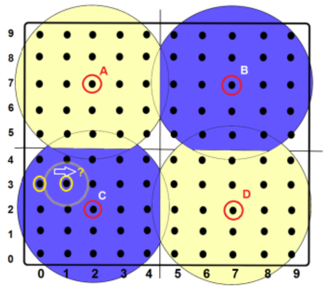
Using bidding function node (1,3) won the auction



STEP 4

Algorithm 3 :
Sensing,transmit

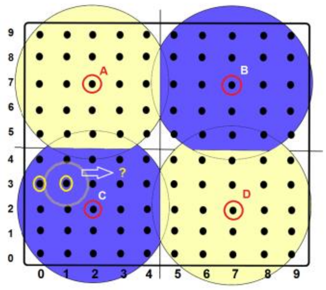
node(1,3) Keep tracking
- **Condition 1: isInRange() ? true**
- **Condition 2: DTT< 1 ? true**



STEP 5

Algorithm 3 :
Sensing,transmit

transmit current stored info and then keep tracking
- **isInRange() ? true**
- **DTT<1, ? false**



STEP 6

Algorithm 3 :
Sensing,transmit

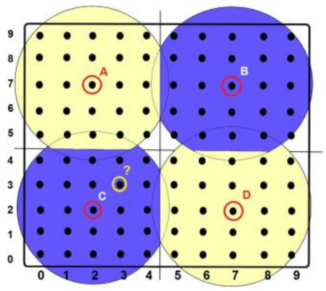
OUT OF RANGE

node(1,3) send signal to **cluster_head** to terminate node(1,3)

cluster_head terminate currently active node(1,3)

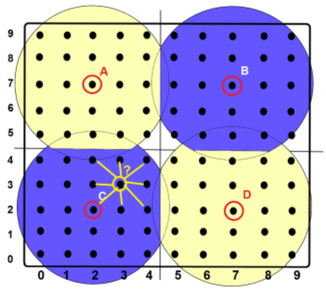
if object in the range of **cluster_head**,select another node

else inform Node C to select another **cluster_head**

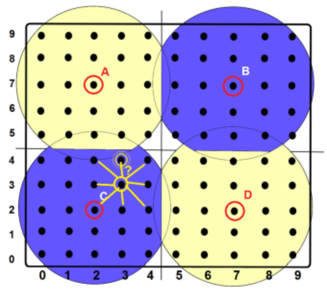


STEP 7

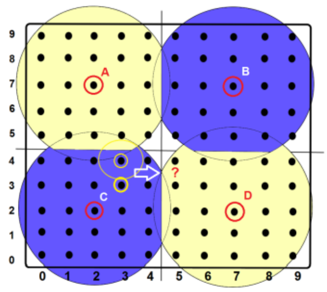
Repeat STEP 3



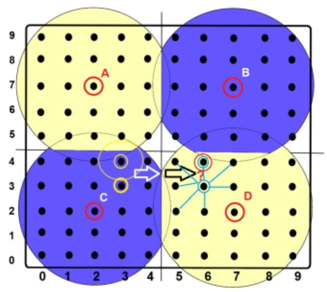
STEP 8



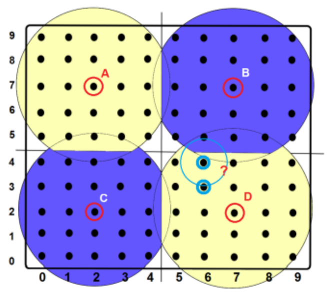
STEP 9



STEP 10



STEP 11



Step 3

STEP 12

We have uploaded the animated demonstration here: **http://bit.ly/WSNBRAC**

V. EXPERIMENTAL RESULTS AND EVALUATION

In this simulation we have taken 100 sensor nodes in the area of 10-meter X 10-meter area. The visibility area of each sensor node is assumed to be 1 meter. One of the main part of this system is the four main sensor nodes which is always be in active mode. Which are powerful than the other 96 nodes. In this system we have divided the whole area into 4 units. Each unit has 24 sensor nodes and 1 main sensor node. 1 main sensor node covers the rest 24 in its area. So each of the main sensor node covers 25% of the total system. When an unwanted substance comes under its (main sensor node) sensing range it calculates and makes the nearest node of the substance as cluster head. So in this system cluster head is not fixed. It is being selected dynamically by the main sensor node.
We consider the following equations for remaining total energy and used energy calculation.

Remaining total energy (Stored energy) =

$$\sum_{\substack{i=node(0,0)\\ i\neq A,B,C,D}}^{i=node(9,9)} \text{RemainingEnergy(i);}$$

Used energy =

initial energy- remaining energy

For simulation we are assuming some events.

In worst case at least one (any one) node will continuously track object and we are also assuming on an average after every 5 minute the transition (changing of currently active node because of changed location of the tracked object) will occur. During transition, at a time maximum of 8 nodes will remain active and between two transitions 2 nodes remain active. One is currently active node and another one is the cluster head.

In this case we are assuming total life time of a node is 10 days (14,400 min). It means one node can be active/live for 10 days' max if standby.

So, for every minute of activation, each node will lose 0.006944% of its initial (100%) charge. And during the transition period 8 nodes(max) remain active simultaneously. This will cost (0. 00007233 × 8) % charge of initial charge of total system.

Total Transition = time÷5; //if transition in every 5 min
One node for 1 min = 0.006944 % //discharge from own
One node for 1 min = (0. 006944 ÷ (100-4))
= 0.00007233% // From total system

Total used energy = energy usage(transition)
+ energy usage(non-transition)

= (Total Transition × cost per transition)
+ (Total Transition × 1 min of system's cost × 2 × 5)

Considering static allocation of energy
Static use of energy = 0.006944% loss/min

Using these calculations and taking transition period of 1-5 minute we draw graphs to compared the scenario where the system consumed all of its energy.

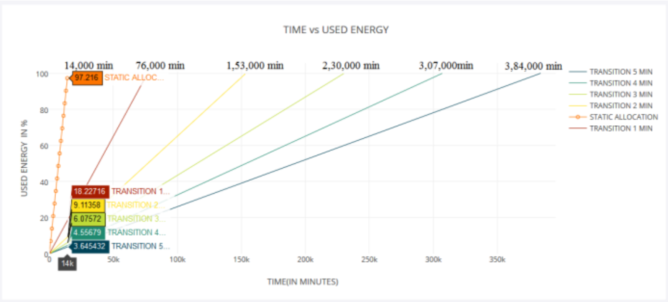


Fig. 3.1: Time vs Remaining energy
(After 14,000 min static allocation consumed total energy)

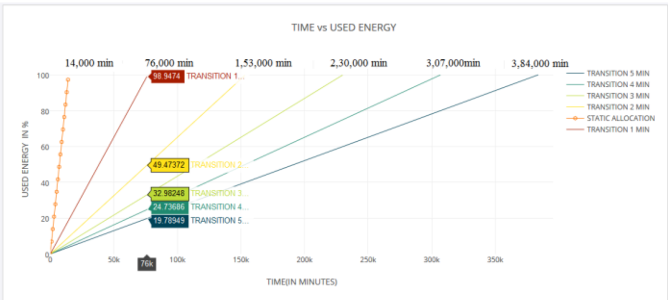


Fig. 3.2: Time vs Remaining energy
(Total energy consumed after 76,000 minutes by 1 min transition)

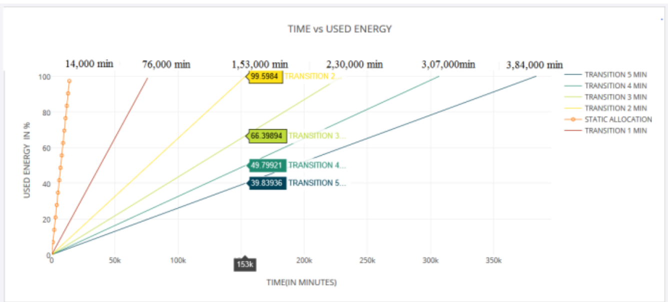


Fig. 3.3: Time vs Remaining energy
(Total energy consumed after 1,53,000 minutes by 2 min transition)

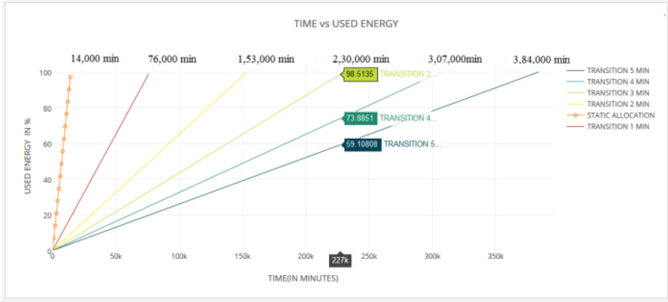


Fig. 3.4: Time vs Remaining energy
(Total energy consumed after 2,30,000 minutes by 3 min transition)

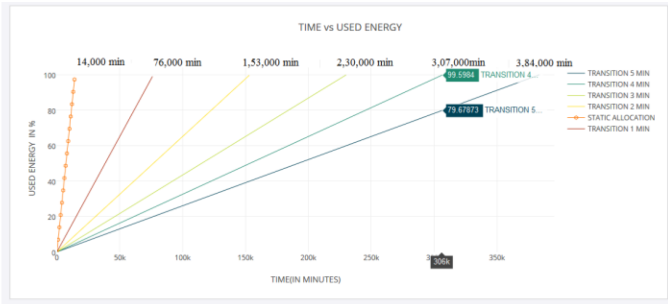


Fig. 3.5: Time vs Remaining energy
(Total energy consumed after 3.07,000 minutes by 4 min transition)

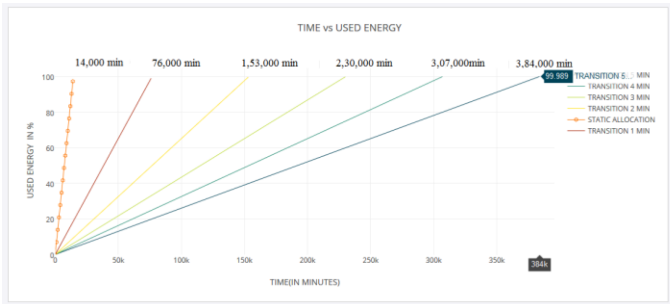


Fig. 3.6: Time vs Remaining energy
(Total energy consumed after 3.84,000 minutes by 5 min transition)

Fig.3.1-3.6 shows remaining energy according to the change of time in worst case scenario. We started time from 0 minute where remaining energy was full (100% initially) as well as consumed energy was 0 and ended at 14000, 76000, 153000, 230000, 307000, 384000 minute respectively for Static Allocation, Transition period of 1 to 5 minutes. We took 387 reading by changing time with 1000-minute interval, until the remaining energy got down to 0% approximately or in other words till total energy was consumed. Fig. 3.1 to 3.6 shows that our approach provides better performance in terms of remaining energy comparing with the existing method. On the other hand, it shows that our method shows better performance in terms of total energy consumption.
According to this simulation result this 100 node can sense the area for 3,84,000 minute (5 min transition) in worst case scenario, where 2 nodes were continuously active and at every 5 minute there were a transition period with 8 active nodes.

VI. CONCLUSION

We have proposed a context aware energy allocation algorithm to maximize the lifetime of the network. In the proposed system we have used three different algorithms to collect data and transmit it when necessary. Our target was to use the Wireless Sensor Network in such a way that we can avoid wastage of energy as much as possible as well as get information. In this paper we have used a composite algorithm to select cluster head and allocation of jobs. Our proposed system focuses on job allocation so that all the nodes do not need to stay awake or active all the time. If we maintain some conditions this system can give very good solution like the main sensor nodes need to be constantly active (by using Solar System or with multiple backup power) and if simultaneously multiple event doesn't happen.

The simulation results show that the system saves significant amount of energy comparatively.

VII. REFERENCES

[1] Lokesh. B. Bhajantri, Nalini. N, Gangadharaiah. S, " Context Aware Resource Allocation in Distributed Sensor Networks ", International Journal of Wireless & Mobile Networks (IJWMN) Vol. 4, No. 2, April 2012 DOI : 10.5121/ijwmn.2012.4203

[2] I.F. Akyildiz, W. Su, Y. Sankarasubramaniam, E. Cayirci, "Wireless sensor networks: a survey", Computer Networks 38 (2002) 393–422, http://www.ece.gatech.edu/research/labs/bwn/surveys/sensornets.pdf

[3] Jinseok Yang, Sameer Tilak, Tajana S. Rosing ,"An Interactive Context-aware Power Man-agement Technique for Optimizing Sensor Network Lifetime", http://seelab.ucsd.edu/~jinsuksfa/Papers/issnip2013.pdf

[4] S. Amit,C. Anantha,"Dynamic Power Management in Wireless Sensor Networks",07407475/01/$10.00, 2001 IEEE, http://www-mtl.mit.edu/researchgroups/icsystems/pubs/journals/2001_sinha_dtc.pdf

[5] J. Vana, "Power Management in Wireless Sensor Networks with High-Consuming Sen-sors", JELICIC, QUALIFYING DOCTORAL EXAMINATION, APRIL 2011, http://www.fer.unizg.hr/_download/repository/VJelicic,KDI.pdf

[6] K. Surender , P. Manish and B. Bharat, "Energy Efficient (EECP) Clustered Protocol for Heterogeneous Wireless Sensor Network", July - 2013, pp. 1448-1453.

[7] Manjusha M S1, K E Kannammal2, "Efficient Cluster Head Selection Method For Wireless Sensor Network", International Journal of Computational Engineering Researc,|Vol, 04,Issue, 2.

[8] Y. Mohamed, Y. Moustafa, A. Khaled, "Energy-aware management for cluster-based sensor networks", Computer Networks 43 (2003) 649–668.

[9] Geoffrey Mainland, David C. Parkes, and Matt Welsh, "Decentralized, Adaptive Resource Allocation for Sensor Networks"

[10] https://www.techopedia.com/definition/25651/wireless-sensor-network-wsn

[11] http://wireless.ictp.it/wp-content/uploads/2012/02/WSN-Applications.pdf

[12] L Anumeha,G A Kulkarni,"A Review of Wireless sensor networks System for monitoring of Asthma Trigger Factors",International Journal of Advanced Research in Computer and Communication Engineering,

Vol. 5, Issue 2, February 2016.

[13] M. Gayathri,I. Harish, "Smart Home Power Management System Using Zigbee",International Journal of Emerging Trends in Engineering and Development, Issue 5, Vol.2, February-March 2015,ISSN 2249-6149, http://www.rspublication.com/ijeted/ijeted_index.htm

[14] Heinzelman W.R, Chandrakasan A., and Balakrishnan H.: "Energy-Efficient Communication Protocol for Wireless Micro sensor Networks".2000.

[15] S.Goel and T. Imielinski, "Prediction-based monitoring in sensor networks: Taking lessons from MPEG", in Wireless Extensions to the Internet Computer Comm. Review, Vol. 31, Oct. 2001, pp. 82–98.

[16] S.Bandyopadhyay, E.J. Coyle, "An Energy Efficient Hierarchical Clustering Algorithm for Wireless Sensor Networks," in: Proceeding of INFOCOM 2003, April 2003.

[17] G.J. Pottie, W.J. Kaiser, Wireless integrated network sensors, Communications of the ACM 43 (5) (2000) 551–558.

[18] I. F. Akyildiz, W.Su, Y. Sankarasubramaniam, and E Cayirci, "A survey on sensor networks", IEEE Communications Magazine, vol. 40, no. 8 pp. 102–114, Aug. 2002

[19] P H. Chou and C. Park, "Energy-efficient platform designs for real world wireless sensing applications", in Proc. Int'l Conf. Computer Aided Design (ICCAD), San Jose, CA, USA, Nov. 2005, pp. 913–920.

A Smart LAN Infrastructure for VoIP Based Wireless Communication

SM Tasdid Swad
Department of Computer Science and Engineering
BRAC University
Dhaka, Bangladesh
tasdid.swad@gmail.com

Md. Rokebul Islam
Department of Computer Science and Engineering
BRAC University
Dhaka, Bangladesh
m.rokebul.islam@gmail.com

Sadia Intesar
Department of Electrical and Electronic Engineering
BRAC University
Dhaka, Bangladesh
sadia.intesar@gmail.com

Jia Uddin
Department of Computer Science and Engineering
BRAC University
Dhaka, Bangladesh
jia.uddin@bracu.ac.bd

***Abstract*—An advanced infrastructure for communication based on smart phone and free of internet connectivity is a time demand. In this paper, we proposed a smart local area network (LAN) infrastructure for voice over internet protocol (VoIP) based wireless communication model, which will be a replacement of official IP phones and conventional intercom system with significant low cost and complexity. The proposed model is a hybrid peer-to-peer (P2P) infrastructure with session initiation protocol (SIP) calling feature into a LAN through wireless access points (WAP). Multiple applications like voice or video call, instant messaging and web based community services along with a resourceful server will be accessible by smart phone's android applications without using internet or cellular networks.**

Keywords-AWB-Codec; Hybrid P2P; IoIP;PPVPN; SIP; WEP

I. INTRODUCTION

The growing demand of globalization necessitates the need for an uninterrupted and secure communication model with attractive user modules. Wireless connectivity is an essential feature of modern communication, for this very reason mobile devices have turned the world into a global information hub. This has given us access to infotainment (information + entertainment), which has become available at a relatively moderate price. In light of growing demand for information it feels like the right time to make communication free of expense while maintaining an enriched user friendly module. Our proposed model is an attempt towards this developing trend of communication technology. It is estimated that every cellphone user in the next two years or sooner will graduate to smartphones, which translates to people more efficiently handling technical aspects like app-to-app calling or choosing software-as-a-service (SaaS) [13]. Our approach was to integrate various existing approaches and focus on improved interface while using minimal hardware support.

Voice over internet protocol (VoIP) is the cheapest method of communication and internet protocol (IP) based infrastructures are operating successfully in different parts of world. Open Garden [2] created a messaging app called 'fire chat' [3] in Singapore, which could send and receive messages without an internet connection. Clear-Com [4], a United States based global company of communication system, provided wireless fidelity (Wi-Fi) based virtual local area network (VLAN) communications in stadiums for the opening and closing ceremonies of the 2010 Winter Olympics and the 2010 FIFA World Cup. In [5], digital acoustic is providing a 'daisy-chained' intercom which promises greater feasibility. Intercom over IP (IoIP) is now used in some countries but not in Bangladesh as of yet. VoIP is comparatively a familiar term used by international gateways (IGWs) and internet service providers (ISPs) which concentrates on profit earning. Wired technology still holds dominance, with 'advanced' IP phones and intercoms prevailing in offices and homes respectively.

We plan for an overhaul of the situation by making the system entirely wireless and accessible by smartphone apps (android based). Our plan is to incorporate numerous community services, such as news headlines, sports update, weather forecast and download utility like pdf, movie or song to be available for users free of cost. Our proposed model has the following features:

- A free app for voice call and instant messaging service.
- A system that stands without internet connection.
- Cloud computing based services for a server system.
- System integration to ensure cost effective manufacture.
- Popular community services with an appealing look.

Development is a continuous process. Let us find other potentials of our proposal which can be well implemented with few incorporation in methodology:

- Replacing old intercom system of home apartments and the IP phones of corporate offices.
- Security or employee monitoring for a large industrial area.
- Digitalization of offices or markets with new features.
- A community based application or a social network site.
- Large Wi-Fi zone like developed countries.

The rest of the paper is organized as follows: Section II describes the problem statement, Section III presents detailed overview of proposed model, detailed implementation approach is in section IV, and finally section V concludes the paper.

II. PROBLEM DESCRIPTION

With each new advancement in communication technology some other means that has been satisfactory before the new technology came into the picture becomes obsolete. Some countries have managed to introduce alternative models of communication into society and simply use it on a regular basis. Bangladesh is lagging behind and the likely causes are difficulty in device installation and inadequate equipment knowledge, maintenance and cost, conflict of benefit and interest, and reluctant and unwilling attitude.

A. Device installation and inadequate equipment knowledge

Advanced infrastructure installment seems hazardous especially in remote sites due to the need of a whole rack of equipment. Common users without prior knowledge of net configuration may not find localized switching equipment satisfactory or user friendly. That is why many users are not interested to go for activation hurdles of new technology..

B. Maintenance and Cost

Due to the obvious hazard regarding installation maintenance is rendered costly. The need for system monitoring may seem expensive, however upon examination we have found it not be be as expensive as is generally estimated.

C. Conflict of benefit and interest

Cellular mobile companies may not welcome an alternative technology that allows communication free of cost. Conventional phone calls use public switched telephone network (PSTN) [14], which will become a redundant. With the boom of low cost smartphones, phone calls using cellular network will cease to exist.

D. Reluctant and Unwilling Attitude

Many initiatives fail in third world countries due to unwillingness towards change. We hardly accept diversification unless it has been imposed or offered for free.

Now let us discuss what we tried to implement in our proposed model. A simple approach to this system can be compared with a common phone feature like ‘SHAREit’ [15] that most of us have possibly used. We are always able to transfer files using Wi-Fi of our laptops or handsets. We do not need to be connected to internet to transfer data. We can also transfer voice over Wi-Fi and it exists in the market. Apps like Skype [17], Viber [16] does not need internet connection if the communicating people are in the same area or under same coverage of Wi-Fi. Generally a Wi-Fi range is limited from a room to a floor but practically it is possible to extend the range to a city or cities beyond. A simple interconnection or device configuration may result to large area Wi-Fi coverage. That is how developed countries first created a zone and later made a link over web server to provide service that we commonly refer to as ‘free internet’ or ‘whole city Wi-Fi’. An infrastructure of this pattern may have its centrally located server too, instead of a link over the internet. The server is capable of storing user data like, i-Cloud [18] does and be able to distribute its storage over Wi-Fi channels, that allows users to do what we call downloads. Our proposed system is also fashioned for areas where the touch of internet is yet to be available. Internet or wide area network (WAN) is nothing but aggregations of interconnected networks [1]. Our netizens (network citizens) will be an internet user who will not have the risk of losing important data sometime when you need those most even if connectivity is lost.

III. PROPOSED MODEL

Figure 1 shows the proposed model of VoIP based smart LAN. This model is a VoIP system where the voice or media packet would transmit through IP.

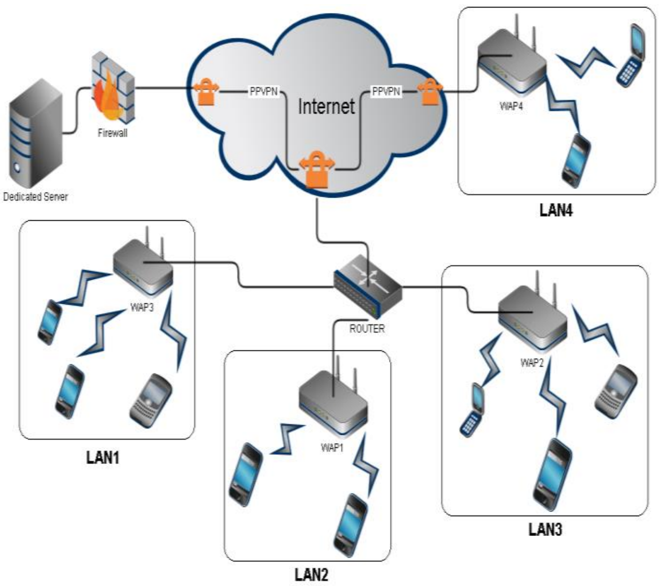


Figure 1. Proposed model layout.

We have used VoIP and IoIP systems but they are fixed in some places, which means the mobility and the ease of use is challenged. IP calling is cheaper than other approaches, so we integrated this idea with Wi-Fi system.

There will be android mobile application which will work as a user agent [20] and a cluster of these user agents will be connected to a gateway, which is actually wireless-access-point (WAP) with the help of Wi-Fi routers. These wireless routers will be connected with a border router. There will be a dedicated server for the users to fetch the login information and some other services. If the system is implemented in a LAN then user should be able to send and receive voice call, messages without an internet connection. This part of our model will be implemented first in small home apartments and will also be marked as a replacement of old intercom systems of those home apartments. They will have a better feature like 'each of your smart phone is a single intercom'. It will be using P2P connections to link other nearby app enabled devices. We may choose to execute any of existing WAP model based on how user or traffic rate demands. Wi-Fi routers with less throughput for less crowded area and high throughput for highly crowded area. Then we should work with the wireless security part of model. Security enhancement of wired equivalent privacy (WEP) Protocol i.e. IEEE802.11b with dynamic key management [12] may be helpful in this regard. We are using IEEE802.11b because it is more economical and easier to get. If better service is the priority, then the 5th generation IEEE802.11ac is available for higher throughput. For long distance communication i.e. outside the LAN, we have to connect the proposed system with the internet. To keep the line secure we could use [21] provider-provisioned-virtual-private-networks (PPVPN) [RFC: 4026].

IV. Implementation of Proposed Model

The IP technology is used widely for all of IoIP products and exists in various developed countries. The benefit of IP based communication is the capability to use available existing IP infrastructure and this is very much implementable to a small extent with field level of technical knowledge. This setup can be divided into three parts: network elements, security, and traffic control.

A. Network Elements

The call is based on Skype or Kaaza architecture [6] and session initiation protocol (SIP) [7]. To create such network system, there are some entities that help this system creating its network. For a simple structure we will be using following network elements: user device, dedicated server, and routers.

- *User Devices:* The network element which will be the interface between the client/user and the system is the smart phone of the targeted user. There will be an android mobile application by which user will connect with the system and call other clients. The working procedure of the application can be arranged in three parts: login and initialization, calling, and storing super nodes.
    - ✓ *Login and initialization:* A client opens a transmission control protocol (TCP) and a user datagram protocol (UDP) [19] listening port at the port number configured by network admin. Client will listen for incoming connections from the ports. At first our software will send a hypertext transport protocol [22] (HTTP) 1.1 GET request to the dedicated server which we will be using to control the users' basic information. The same server will be used to authenticate user name and password of the client. This should lessen the CPU and memory usage of a smartphone, which makes the prescribed system an effective one. Location of the server will be provided by the network admin. As the system gets bigger, we will use a dynamic host control protocol (DHCP) server. After authentication, the client will advertise its existence to the friends of the client by sending a broadcast packet with a media access control (MAC) address of FF: FF: FF: FF: FF: FF. Each user will have a dial code, so the friend of our experimental user (EU) will accept the packet and show the EU is online. The friend of the EU will again send a control packet advertising that they are online. Those who are not friends, they will simply ignore the broadcast. Figure 2 shows a simple illustration of initialization of the system.

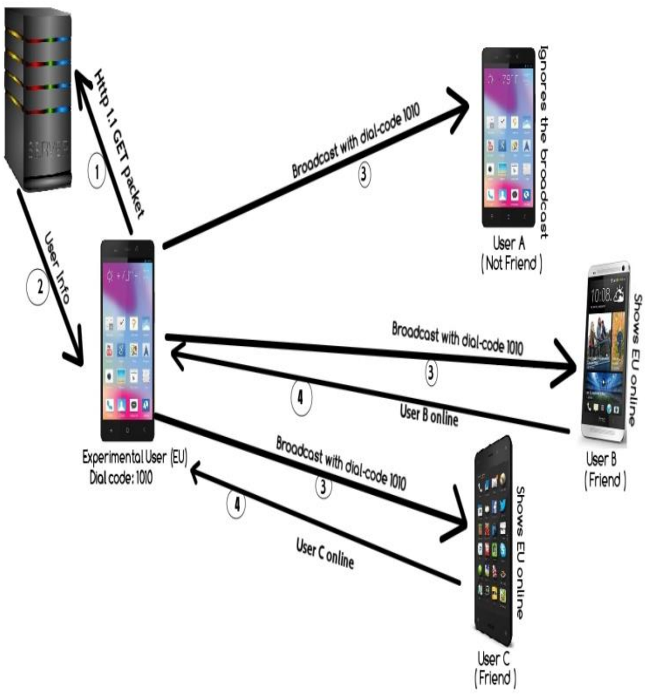


Figure 2. Simple illustration of initialization of the system.

    - ✓ *Calling:* There is no use of this system if the call quality is compromised and if the system is half-duplex. We will use the adaptive multi-rate wideband codec G.772 [8], whose sample rate is 16 kHz using 14 bits providing an effective pass-band of 50 to 6400Hz. G.772 allows maintaining improved call quality than G.711 at an available bandwidth of 32Kbps. We will be using TCP for controlling and signaling the data bits for its reliability and for voice data both UDP (RTP) and TCP. The proposed system is a hybrid P2P. Once the connection between caller and callee is established, there will be no need of a server. The process will be executing on its own. Figure 3 illustrates a basic call session setup between two parties: caller and callee. As depicted in Figure 3, initially the caller sends an INVITE [7] request packet to the callee. Caller waits for the 180 RINGING [7] packet (provisional response). If caller does not get the 180 that is RINGING packet within the timeout session, it automatically understands that the INVITE packet is lost. So, the caller sends the INVITE packet again. Whenever callee gets the INVITE packet, it sends the 180 RINGING packet. After the call is being picked, callee sends a 200 OK packet. Caller sends an ACK packet afterward.

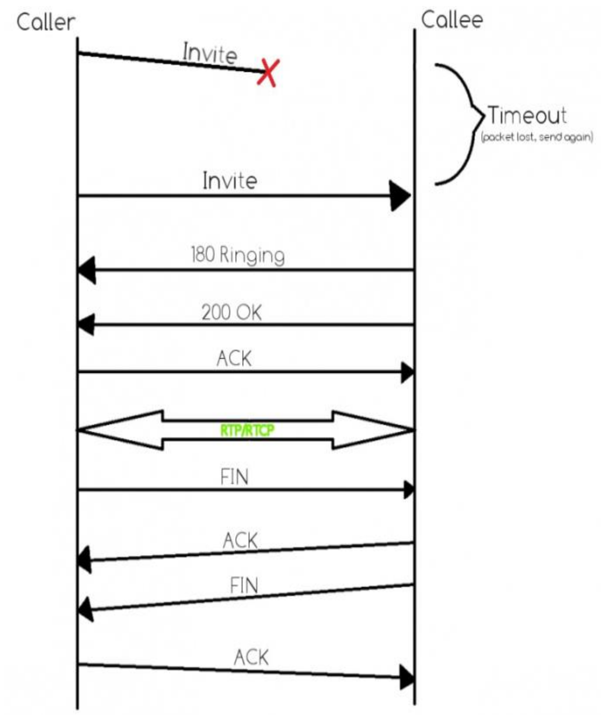


Figure 3. Basic call session setup [7].

After receiving the ACK packet, callee initiates a real time transport protocol (RTP). RTP [RFC 3550] will use for its rapid response. It basically runs over the UDP protocol and is used for media streaming. The user's voice which we are trying to send over IP is nothing but a stream of media. A buffer would collect the audio stream which is user's voice and it goes to the RTP library [9]. The stream is multiplexed and encoded in RTP packet by the RTP library. Then with the RTP payload the other layers of the packet UDP, IP, and Ethernet are added, respectively. After the conversation, as soon as caller hangs up; an FIN packet is sent to the callee. Just like a simple http session, this session will be terminated.

- ✓ *Storing Super Node:* Super node is an advanced criterion for load balancing of the network. For a gigantic network, super nodes reduce the extra packets, search time [11] and ultimately the congestion of the network. Like Skype we will use the hybrid p2p model where every client stores a list of super node in its cache (host cache). A super node is an ordinary host's end-point on our system's network. Any mobile device i.e. android operated smart phone with adequate CPU, memory, and network bandwidth will be eligible to become a candidate for being a super node. The best match for these criteria will be considered to be one. When the network will be big enough, an ordinary node that is not eligible to become a super node; must connect to a super node while initiating. After getting connected with a super node, the client will authenticate itself with the dedicated server. The server ensures that login names are unique across the system's name space. As smart phones are not capable of storing moderate amount of data for just an application so we will be using only a dedicated server, not the super nodes. But with the advancement of technology and increasing capability of a smartphone, soon we will shift to the super node technology.

- *Dedicated server:* For the next step we are going to need a server for storing all the information of the clients i.e. friend list, chat history, call history, and login information. Primarily for home intercom system, we are not looking at hybrid P2P. But for better service and to accommodate more users we ought to consider hybrid P2P system. For a small LAN, the server will work like a router. All the switching will be done within the server. But for larger model which consists of two or more LANs, we will use dedicated border router(s). For a small network, such as home intercom service, the server will store the dynamic host configuration protocol (DHCP) information as well as the login information. For a huge network, only one central DHCP server will cause more congestion as a result, the server will be down. So, for such network, there will be dedicated DHCP server, configured in a cisco router that will provide user the IP addresses, subnet address, lease time, border router or the default gateway address, DNS server addresses, WINS server addresses, and domain name. Third party firewall will give the server necessary protection.

- *Routers:* The cluster of user devices will connect with the network through the WAP. Every single WAP will be for a definite amount of the user device because of the wireless throughput of the WAP. We need at least 32 Kbps bandwidth allocated for a single call. So we need to check the client getting connected with the network. When a better WAP i.e. IEEE802.11ac is available more clients can accommodate. If not then the amount of client will be checked and the system will not allow any other client to login. The system will be designed according to the users' preferences. For communication between more than one LAN, a border router needs to be introduced. Network Address Translation (NAT) will be done by the border router(s). These border routers will then route all the RTP packets throughout the network.

B. Security

Security of the system can be divided into two parts, such as- wireless (WEP), and wired (PPVPN).

- *Wireless (WEP):* Wireless security is applicable for the wireless routers or the WAPs. In [12], they have redesigned the system with the dynamic key management. In the conventional systems, static WEP key i.e. the same key is used in encryption so it was easy for the attacker to break in the system. The size of the initialization vector (IV) is 24, which repeats after every 4.07 hours. Once the IV is repeated, attacker can easily extract the information. Dynamic key with 128 bits length increases the repetition time into staggering $2.13*10^{31}$ hours. This makes the encryption almost unbreakable because the dynamic key is depended on the future IV and the previous shared keys.

- *Wired (PPVPN):* To get seamless communication between two or more geographically separated LAN, we have to deploy the PPVPN for tunneling our data through the internet. After connecting the system with the internet, the clients can communicate easily and securely despite of the fact long distance. But to communicate through the internet, there are some risks. Encryption of PPVPN can secure client data and information. PPVPN allows client-edge-device to client-edge-device (CE-CE) P2P connection. PPVPN can be categorized mainly into two parts; L2PPVPN and L3PPVPN [12], L3PPVPN is the better solution for its scalability, management, higher number of subscriber and reduced customer-premises-equipment (CPE) complexity and processing. Table 1 shows the comparison between the two methods.

C. Traffic control: Peer churning rate

Arrival and departure of a peer in the network causes the update process of the P2P overlay [10]. If no packet is sent within the time period, i.e. longer than the set idle time, the peer is considered to be idle (or dead). The peer is considered to be alive when the set ideal time is greater than the no activity time. So if the ideal time is increased, it will lessen the strain of our P2P system of our proposed model.

Table 1: Comparison between L2PPVPN & L3PPVPN [12]

Decisive Factor	*L2VPN*	*L3VPN*
Easy to configure	✓	✓
Scalable		✓
Can accommodate high no. of client		✓
Service Provider management		✓
New service provisioning		✓
Operational with existing ATM/FR VPNs		✓
Many VPN sites		✓
MPLS and BGP installed		✓
Less CPE complexity	✓	✓
Less CPE processing	✓	✓

V. CONCLUSION

In this paper we have designed a model both to experience change and benefit the process of digitalization in our country. We have designed infrastructure and examined possibilities of a communication system which can be accommodated wirelessly both at home and outside. We have shown that a system integrated communication model has all the attributes of reducing anxieties in case of connectionless modern life (be it mobile coverage or internet) and able to manage things better than concurrent system. The cost effectiveness and user friendly module demands it to become a unique project and popular society tool. Our model is aimed to build a society of better living standard by a more advanced technology.

ACKNOWLEDGMENT

Authors are grateful to BRAC University for their constant support especially to CSE department for laboratory facility and EEE department for supplying the necessary equipment's throughout our experimentation and research till to the end stage of development.

REFERENCES

[1] J.F. Kurose, and K.W. Ross, Computer Networks and the Internet, in Computer Networking A Top-Down Approach, 6th Ed, NJ: Pearson Education Inc., 2013, pp. 1-864.

[2] [online]. http://www.intelligenthq.com/social-media-posts/open-garden-sharing-internet/ [Access time: 12 November 2015]

[3] Open Garden, Firechat, [online]. http://opengarden.com/about-firechat [Access time: 03 December 2015]

[4] [online]. http://www.clearcom.com/product/ip-communications [Access time: 01 December 2015]

[5] [online]. http://www.digitalacoustics.com/media_press_ii3mx.htm [Access time: 26 October 2015]

[6] S.A. Baset and H. Schulzrinne, "An analysis of the Skype peer-to-peer internet telephony protocol," in IEEE Infocom'06, Barcelona, Spain, Apr. 2006, pp. 1-12.

[7] J. Rosenberg, H. Schulzrinne, G. Camarillo, A.R. Johnston, J. Peterson, R. Sparks, M. Handley, and E. Schooler. SIP: session initiation protocol.RFC 3261, IETF, June 2002.

[8] International telecommunication union: Recommendation G.722.2, [online]. http://www.seas.ucla.edu/spapl/codecs/G722-2.pdf [Access time: 23 October 2015]

[9] A.S. TanenBaum, The Transport Layer, in Computer Networks, 4th Ed, NJ: Prentice Hall, 2003, pp. 404-406.

[10] D. Stutzbach and R. Rejaie, "Understanding Churn in Peer-to-Peer Networks," in Proc. of IMC'06, Rio de Janeiro, Brazil, pp. 1-13.

[11] D. Hadaller, K. Regan, T. Russell, "The Necessity of Supernodes, pp. 1-25," [online]. http://www.cs.toronto.edu/~kmregan/files/proj_supernode_paper.pdf [Access time: 01 November 2015].

[12] A.B. Mack, "VPNs: The Race Is On" in Proc. International Engineering Consortium on VoIP and Enhanced IP Communications Services, Comprehensive Report series, 2005, pp. 132-137.

[13] R. Rai, G. Sahoo and S. Mehfuz, "Securing software as a service model of cloud computing: Issues and solutions", 2013.

[14] A. Voit, Eric. "Internet Phone To PSTN Cellular/PCS System". 2001: n. pag. Print.

[15] SHAREit FAQ. [online]. http://shareit.lenovo.com/faq.html

[16] Viber. Viber Media (Rakuten Inc.), 2015.

[17] Skype FAQ. [online]. http://www.skype.com/help_faq.html

[18] iCloud, Icloud.com, 2016. [online]. https://www.icloud.com/. [Accessed: 04 May 2016].

[19] W. Buchanan, Mastering networks. Basingstoke: Macmillan, 1999.

[20] W. Bauer, German. "Interface For User/Agent Interaction". 1999: n. pag. Print.

[21] R. Callon and M. Suzuki, A framework for layer 3 provider-provisioned virtual private networks (PPVPNs), 2005.

[22] R. Fielding, J. Gettys, J. Mogul, H. Frystyk, L. Masinter, P. Leach, T. Berners-Lee. HTTP: hyper text transfer protocol. RFC 2616, IETF, June 1999.

[23] Lessard, J., Kessler, G.C., "Android Forensics: Simplifying Cell Phone Examinations," Journal of Small Scale Digital Device Forensics, vol. 4, no. 1, pp. 1–12, 2010.

Design of a Vision Based Person Following Robot

Rakibul Hasan[1],Jahedul Islam[2],Rashed Md. Murad Hasan[3] and Adhri Nandini Paul[4]
[1, 2, 3] Department of Electrical and Electronic Engineering
[4] Department of Computer Science and Engineering
Chittagong university of Engineering and technology
Chittagong-4349, Bangladesh
Rakibulhasan1992@hotmail.com[1],bbadsha60@yahoo.com[2],murad_eee@cuet.ac.bd[3],adhri.nandini@yahoo.com[4]

***Abstract*— For a domestic robot "Person Tracking" is a mandatory feature. There are, in general, two steps to be completed: detecting the target person and following him. Hence, the robot needs applicable algorithm or specific sensor. In this project features detection and following technology is based on Image processing technique. Color detection algorithm is used for person following in domestic environments. The vision system of robot obtains very good features of the target person through the webcam and the exact information about the position of the target person in the environment. The proposed method is programmed in high speed hardware system and using small zone person following method in order to meet the real time requirements.**

Keywords*— Image Acquisition tool; Image Processing; Optical character recognition; RGB Image and Arduino IO tool.

I. Introduction

For domestic robot's improvement, exact human detection and tracking has been extremely active research sector for last few years. The importance of this area arises from its numerous applications such as video surveillance, smart vehicle [1]. One of the broadest application areas is robot vision. The importance of robot vision is that it will enable robots to perceive the external world in order to perform a large range of tasks such as navigation, object recognition and categorization, surveillance [2]. Possible applications for such robots cover assistance for elderly people, carrying bags, luggage and suitcases in airport or shopping centers. Old and sick people can use these for carrying their medicines, portable Oxygen concentrator etc. as they exercise outdoor. With further modification, this kind of system can also be used as babysitter or helping hand for the disabled people. This robot can work in a team helping a human co-worker with transportation of different objects, such as investigation of a hazardous environment. The main requirement for service robots is the ability to detect humans and interact with them in a non-technical, natural fashion.

In this project, the aim is to design a system in such a way that it can detect selected person and follow him/her. Image processing technique has been used to achieve this objective. Person following robots developed until now use several types of ways for tracking a target person and some of them use various sensors like ultrasonic sensor, laser sensors etc. [3 4]. But most of the cases these sensors are very costly. Moreover, ultrasonic sensors have distance limitations, loud noises and air hoses problems and environmental effect problems (temperature, humidity, pressure etc.). Tracking systems with laser range scanners have the problems of noise. Beside this, some authors developed sensors, which are however, inadequate in some applications [5]. In order to avoid the complex data fusion algorithms and human contact sensors, a number of research have been done on vision as the only sensor.

II. Proposed Method

To carry out our aim, to design an artificial vision system to follow a person, the most cardinal step is to build up a platform. In fig.01, the overall designed platform is given.

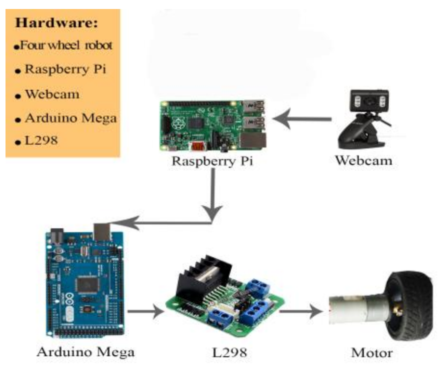


Fig. 1. Basic Design

The detection system is based on Optical Character Recognition (OCR) and tracking system is based on Region of Interest (ROI) algorithm - the red circle of the person's back. Raspberry Pi is a tiny, low-cost, single-board computer designed for teaching. In Raspberry Pi, High level languages such as Python, Java or assembly language are used for coding and debugging. . Laptop was used instead of Raspberry Pi for simplicity. MATLAB is used as central control

III. Methodology

The complete work of the system can be divided into two parts. Firstly, the hardware part of the robot and second, the software part for detection, determining position and implementing logics to generate control commands for the robot's motor. The distance between the robot and the person is always maintained constant.

A. *Hardware Section:*

The webcam was mounted on the Robot and it was connected to the USB port of the Laptop. The specifications of the camera are:

- plug and play USB connection
- Video data format: 24 bit RGB
- 30 fps max
- Resolution 640 x 480 pixels

As the project work is carried out aiming to design an artificial vision system to track a person by OCR and then detect and follow the red circle of the person's back, a large amount of data is produced when webcam is used as a Sensor. The central controller should able to process these huge data at a time. MATLAB being a powerful central controller is able to process these types of high volume data at ease.

Arduino IO tool has been used for interfacing Arduino MEGA 2560 with MATLAB. Arduino Mega2560 controlled the movements of robot. To drive the two DC motors, the motor driver IC L298 was used. Hardware schematic diagram is given below:

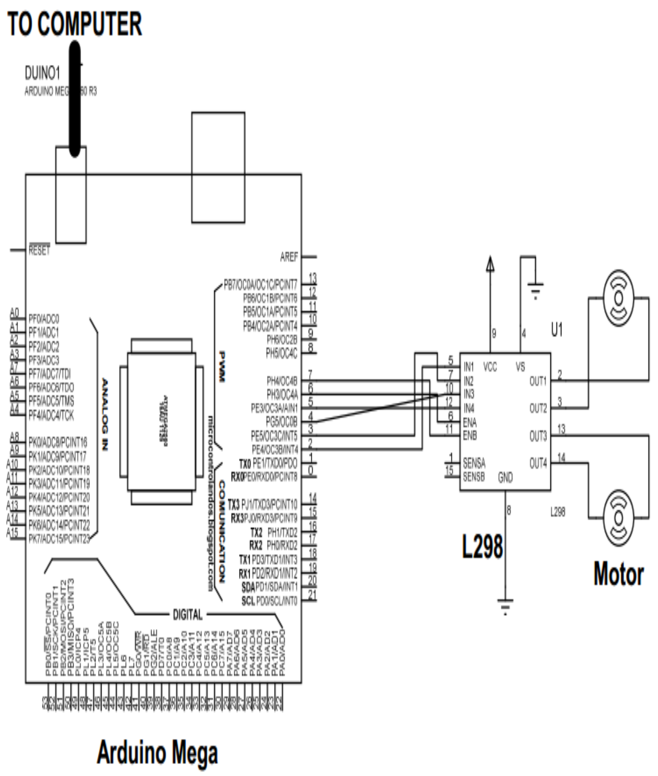


Fig. 2. Hardware schematic diagram.

B. *Software Section*

The Program flow of the project describes how and when which type of work was done.The flowchart for the whole software section is given below:

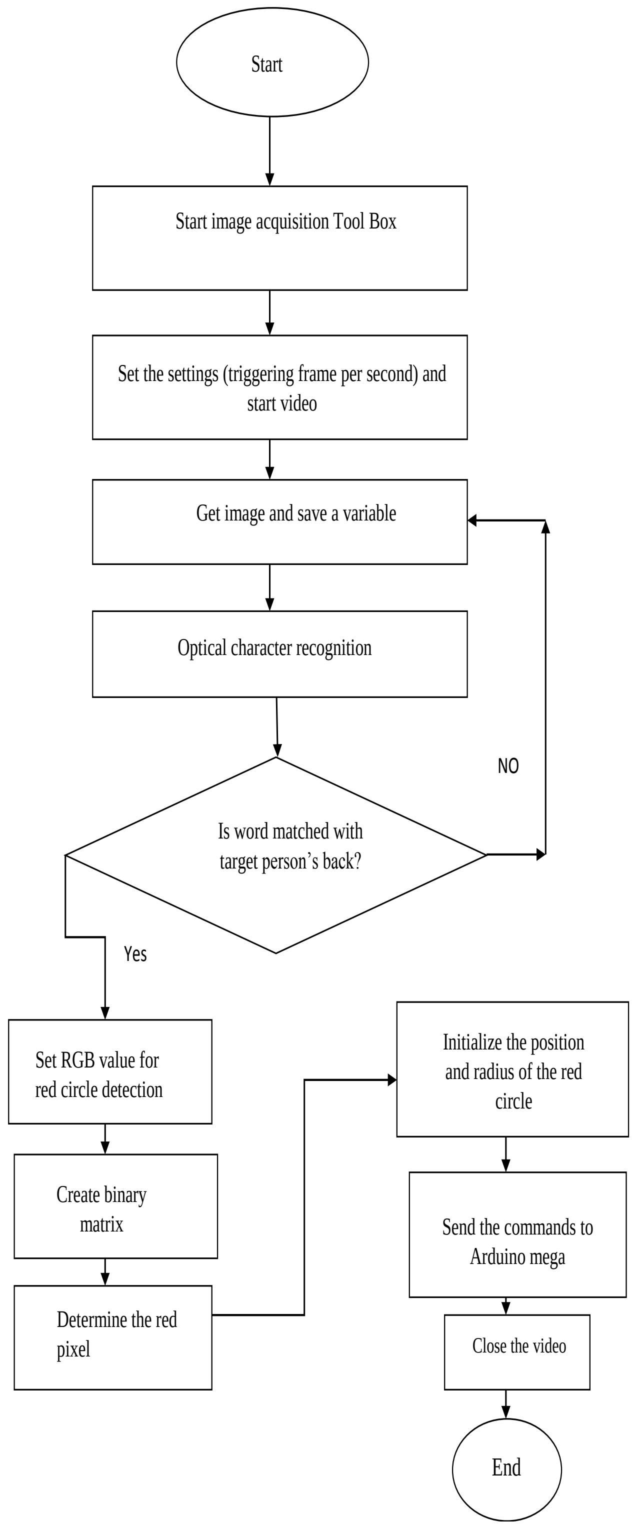


Fig. 3. Flow chart of the program.

1) Image acquisition Toolbox-MATLAB: Image processing Toolbox is an attractive feature of MATLAB, which has a strong and efficient quality to handle images. MATLAB has built-in adaptors to access a web cam. With MATLAB acquisition tool, the control starts capturing the frames. The frames captured are stored in the memory. The setup has been set such way that only one frame is captured per trigger. The number of times for which the trigger can be repeated is infinite.

Fig. 4. Stored picture from webcam.

2) Optical character recognition (OCR): Optical character recognition (OCR) is a process of converting optically scanned bitmaps of printed or written text characters into character codes, such as ASCII. OCR is the most cost-effective and speedy method available to digitize printed texts so that it can be electronically edited, searched, cached more compactly and applied in machine processes such as machine translation, text-to-speech [6]. We used optical character recognition as our detecting method. With OCR, the system will automatically find the same person in adjacent frames from a real time video sequence. Three major steps of OCR are:-

- Processing the image
- Extracting the characters
- Character matching.

a) Processing the image: First the image from the webcam was stored in an array variable using matlab function getsnapshot. Then all data from the data acquisition engine was removed and the samples available property was reset to zero. Once converted to a B & W image, it was found that apart from the region of interest, there are also some unwanted white regions in the image, called 'noise'. The noisy parts of the images have been filtered with Median Filter. The image has been resized, enhanced and the edges were brightened for better detection. Morphological closing tends to smooth the contours of objects and what more joining narrow breaks and filling long, thin gulfs and holes smaller than the structuring element.

*b) Extracting the character:*The image contained components of various pixel ranges. The lower pixel components were removed from the image and measured the length of the components and sorted out the components having approximately same size. Then the characters were separated from the rest of the image, by taking the value of indices among the bounding boxes.

c) Character matching: Each character was separately segmented and scaled down to a standard size [24*42]. Then each character was assimilated with a set of pre-defined character templates. In MATLAB corr2 function was used to find the co-relation factor. The character which gives maximum score was set to select. Then selected character's ASCII values were compared with desired value which is in the target person's back. If the value matches, the target person is detected.

3) Red Circle Detection: The notion of color detection is a part of image processing that involves differentiation among objects based on their color. Any image in RGB format has each pixel in it possessing a set of values for each of the 3 channels. If the range of RGB values for a particular colour can be determined, we want to detect, then while processing the image, the device will search only for those pixels which have RGB values in the range of selected RGB values.

a) Setting the RGB threshold values: The red circle is our main interest. So it is necessary to set the threshold values of the images. The steps for setting the RGB threshold values of the red circle are-

- About 5 snaps were taken at various angles.
- Then each image was read in MATLAB environment.
- Each image was displayed using 'imview' function.
- After activating the 'cursor', the pixel values of the red circle were noted down by placing the pointer at various portions of the circle. The threshold for red component should be the least value of the red component found in the red circle. The threshold for green and blue components should be maximum value of the component found in the circle.

A B&W image array 'I' can be found. 'I' is the result of logically 'ANDed' image matrices fR, fG, and fB. If the three conditions are satisfied, the pixel value of the 'I'is set to '1'.

b) Process of following red circle: The frame captured by the webcam was divided into three parts: 'X1', 'X2' and 'X3' on X-axis.

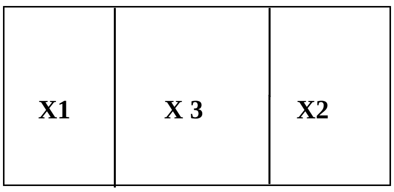


Fig. 5. Plotting of a captured image

$$X1=x/2-radx \tag{1}$$
$$X2=x/2+radx \tag{2}$$

‘radx’ is a variable number which value is dependant on the size of the red circle. It was taken radx=120 in this program. Then co-ordinates of the centre of the frame was calculated (which is x/2,y/2). ‘x’ is the maximum dimension of X-axis (640) and ‘y’ is the maximum dimension of Y-axis(480).

Conditions for tracking the red circle are:

- If the circle is in region X2, it is at the right of the frame. The robot should move right.
- If the circle is in region X1, it is at the left of the frame. The motor should move left.
- If the circle is in region X3, it is at the center of the frame. Then we should use MATLAB `imfindcircles` tool to determine the radius of the circle and compare it with a declared radius. The comparison gives the direction whether to move forward or backward.

The result of Equation (3) provides the direction of the robot: d= ((cbar>=x1)*2 + (cbar<= x2)) (3)

‘cbar’ is the mean of columns of the array derived from the Label matrix using MATLAB find tool. Returning as an array of class double, the Label matrix has the same size as the input image.

4) MATLAB- Arduino Interfacing: Only MATLAB was used for controlling Arduino mega instead of using Arduino software. Arduino IO [7] tool was used for establishing the interfacing between MATLAB and Arduino MEGA. Arduino PWM output with the analogWrite() function was used to control the motor speed.

IV. Result Analysis

The device works in any environment. We just have to adjust the color settings according to the red circle. We tested our robot at our hostel corridor and laboratory and it worked perfectly.

Fig. 6. Practical Implementation.

After Optical Character Recognition the following images were found:

Fig 07: After Image Enhancement

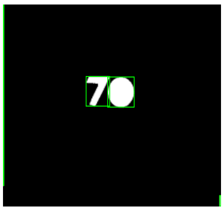


Fig 08: After removing lower pixels

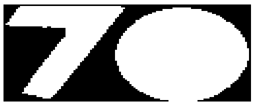


Fig 09: Extracted characters

Selected character’s ASCII values were matched with desired value which is in the target person’s back. After detecting the person, the red circle delivers the information about the position of the person. According to position of circle in the frame, the following images were found:

Fig. 10. Circle is in X2

Fig. 11. Circle is in X1

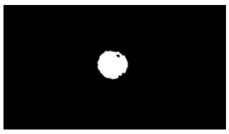


Fig. 12. Circle is in X3

To keep the distance to the person constant, the robot measures the distance of the person by comparing the circle radius and then controls its speed. When the target person moves forward, it moves forward, and when the person stops, the robot moves to a point behind the person and also stops. If the person approaches the robot too much, it backs off.

V. CONCLUSION

In this paper, a person following robot has been introduced which is capable of following a person using only a webcam. It has been tested in the dummy industrial environment developed in the laboratory. The system's accuracy is quite viable. There is a scope to add Artificial Intelligence (AI) [8] in this system which will automatically adjust the color settings at different environments. Furthermore, modified Kalman filter can be applied to solve the problem of multiple objects of same colour properties in the same frame. Moreover, vision quality can be improved by using our android phone as wireless webcam, which will be connected via WIFI. It can also be used as a monitoring system. Video can be transmitted via internet, so one can see the live streaming and monitor the target person.

REFERENCES

[1] Moeslund T B, Hilton A, Krüger V, Sigal L (Ed.) (2011) "Visual Analysis of Humans - Looking at People". Springer-Verlag London.

[2] Ales Ude "Robot Vision", InTech, March 01,2010

[3] Xiaoping Yun, Chuan-Hao Yang, Marcello Romano "A Parson-Tracking Mobile Robot Using An Ultrasonic Positioning System". Monterey, California. 2005.

[4] Takafumi S, Takashi Y, Manabu N, Hideichi N, Seiji T, Nobuto M (2008) "Person Following Robot with Vision-based and Sensor Fusion Tracking Algorithm". Computer Vision, Book edited by: Xiong Zhihui, pp. 538, I-Tech, Vienna, Austria.

[5] Tim Braun , Kristof Szentpetery´ , Karsten Berns (2011) Detecting and Following Humans with a Mobile Robot. University Kaiserslautern, Robotic Systems Group, Kaiserslautern, 67653, Germany.

[6] Ravina Mithe, Supriya Indalkar, Nilam Divekar"Optical Character Recognition" InternationalJournal of Recent Technology and Engineering (IJRTE), Volume-2, Issue-1, March 2013.

[7] http://www.mathworks.com/hardware-support /arduino-matlab.html

[8] Brunette, E.S. ; Sch. of Eng. & Adv. Technol., Massey Univ., Palmerston North ; Flemmer, intelligence" 4th International Conference on Autonomous Robots and Agents, (ICARA), Pp 385-392 , February 2009.

:

An Expert System for Clinical Risk Assessment of Polycystic Ovary Syndrome under Uncertainty

Rezaul Karim[1], Md. Khaliluzzaman[2]*, Tanzim Mahmud[3] and Deepak Kumar chy.[4]

[1]Dept. of Computer Science and Engineering, University of Chittagong, Chittagong-4331, Bangladesh
[2*]Dept. of Computer Science and Engineering, International Islamic University Chittagong, Chittagong-4203, Bangladesh
[3]Dept. of Computer Science and Engineering, Textile Engineering College, Begumgonj, Noakhali-3821, Bangladesh
[4] Dept. of Electrical & Electronics Engineering, University of Information Technology & Sciences, Dhaka-1212, Bangladesh
pinnacle_of_success@yahoo.com, *khalil_021@yahoo.co.in, tanjim_cse@yahoo.com, dk_chy53@yahoo.de

***Abstract* - This paper describes a prototype clinical expert system for risk stratification of patients with polycystic ovary syndrome (PCOS). Polycystic ovary syndrome (PCOS) is the most common hormonal disorder among women of reproductive age. It is a heterogeneous disorder of uncertain cause. Since the symptoms of PCOS are seemingly unrelated to one another the condition is often overlooked and undiagnosed. The determination of accurate degree or intensity of PCOS signs is difficult for the physician. Hence, the accuracy of the diagnostic process is difficult to achieve. The signs and symptoms of PCOS are usually expressed in qualitative and quantitative ways. Since the qualitative factors cannot measure in a quantitative way, various types of uncertainties may occur, such as incompleteness, vagueness and imprecision. For that, it is necessary to address the issue of uncertainty by using appropriate methodology. However, no existing system is able to address this issue of uncertainty. Therefore, this paper demonstrates the application of a novel method, named belief rule-based inference methodology-RIMER; this prototype can deal with uncertainties in both clinical domain knowledge and clinical data. This paper reports the development of a Belief Rule Based Expert System (BRBES) using RIMER approach, which is capable of detecting the PCOS by taking account of signs and symptoms.**

Keywords: Belief Rule Base Expert System (BRBES), Uncertainty, RIMER, Evidential Reasoning, Polycystic ovary syndrome (PCOS), Signs and Symptoms

I. INTRODUCTION

PCOS is a common complex condition in women associated with reproductive and metabolic features. It is a chronic disease with manifestations across the lifespan and represents a major health and economic burden. A common ovulation problem that affects about 5% to 10% of women in their reproductive years is polycystic ovary syndrome (PCOS). PCOS is a hormonal imbalance that can make the ovaries stop working normally. In most cases, the ovaries become enlarged and appear covered with tiny, fluid-filled cysts. Infertility is one of the most common PCOS symptoms. Because the symptoms of PCOS are seemingly unrelated to one another, the condition is often overlooked and undiagnosed. The determination of accurate degree or intensity of PCOS signs is difficult for the physician. Hence, there is a risk of having incomplete information to reach a conclusion through PCOS diagnosis. Therefore, the accuracy of diagnostic process is difficult to achieve. Since PCOS symptoms are subjective in nature, it inherits uncertainty. Therefore, it can be seen that the determination of signs and symptoms in a quantitative way is difficult to achieve. The traditional system is used in PCOS diagnosis to detect the disease. But this system is not working in uncertain. So, ultimately the accuracy in disease detection process is hampered. As human life is directly involved with the medical diagnosis process, disease diagnosis process accuracy is very important factor for saving human life.

There is no specific test to definitively diagnose polycystic ovary syndrome. The diagnosis is one of exclusion, which means doctor considers all signs and symptoms and then rules out other possible disorders. Uncertainty exists in almost every stage of a diagnostic process. Sources of uncertainties may include that patients can't describe exactly what has happened to them or how they feel, doctors and nurses cannot tell exactly what they observe, and laboratory report results may be with some degrees of error. Physiologists do not precisely understand how the human body works, medical researchers can't precisely characterize how diseases alter the normal functioning of the body, pharmacologists do not fully understand the mechanisms accounting for the effectiveness of drugs, and no one can precisely determine one's prognosis.

Researchers and scientists have built and applied various methods in this growing research field. For uncertainty selection RIMER is treated as an appropriate method to solve the certain problems [13][14]. In [7], ER deals with problems under various uncertainties such as incomplete information, vagueness, and ambiguity consisting of both quantitative and qualitative criteria. In particular utility theory the ER approach is developed based on decision theory [1][11], artificial intelligence in particular the theory of evidence [9][10]. A belief structure is used to model a judgment with uncertainty. some linguistic referential value such as excellent, average, good and bad are used to evaluate qualitative attribute such as location or safety [20][21]. In this way, the issue of uncertainty can be addressed more accurately and robustly during decision made. The belief rule based inference methodology-RIMER [15] has addressed such issue by proposing a belief structure which assigns degree of belief in the various referential values of the attributes.

Consequently, traditional diagnosis, carried out by a physician, is unable to deliver desired accuracy. Moreover, traditional PCOS diagnosis is time consuming and costly. Hence, this paper presents the design, development and application of an expert system that will diagnose PCOS precisely in a short time with low cost.

This paper is organized as follows. In Section II briefly described the belief rule base inference methodology-RIMER. In Section III demonstrated the application of BRB to diagnose PCOS. In the next Section results and achievements are represented. Finally, the paper is concluded in Section IV.

II. RIMER TO DEVELOP BRBES

In RIMER, Belief Rule Base (BRB) can capture complicated nonlinear causal relationships between antecedent attributes and consequents, which are not possible in traditional IF-THEN rules. BRB is used to model domain specific knowledge under uncertainty, and the ER approach is employed to facilitate inference. This section introduces BRB as a knowledge representation schema under uncertainty as well as inference procedures of RIMER.

A. Modeling domain knowledge using BRB

Belief Rules are the key constituents of a BRB, which include belief degree. This is 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 [15][24]. The knowledge representation parameters are rule weights, attribute weights and belief degrees in consequent attribute, which are not available in traditional IF-THEN rules. A belief rule can be defined in the following way.

$$R_k: \begin{cases} IF\ (p_1 is\ A_1^k) \cap (p_2 is\ A_2^k) \cap \ldots\ldots\ldots\ (p_{T_k} is\ A_{T_k}^k) \\ THEN\ \{(C_1,\beta_{1k}), (C_2,\beta_{2k}), \ldots\ldots\ldots, (C_N,\beta_{Nk})\} \end{cases} \quad (1)$$

$R_k: (\beta_{jk} \geq 0, \sum_{j=1}^{N} \beta_{jk} \leq 1)$ with a rule weight θ_k attribute.

Weights $\delta_{k1}, \delta_{k2}, \delta_{k3}, \ldots\ldots, \delta_{kT_k}, k\epsilon\ \{1, \ldots\ldots, L\}$

Where p_1, p_2, p_3 $\ldots p_{T_k}$ represents the antecedent attributes in the k^{th} rule. $A_i^k(i = 1, \ldots\ldots, T_k, k = 1, \ldots\ldots, L)$ represents one of the referential values of the ***i***th antecedent attribute P_i in the k^{th} rule. C_j is one of the consequent reference values of the belief rule.

$\beta_{jk}(j = 1, \ldots\ldots, N\ k = 1, \ldots\ldots, L)$ is one of the the belief degrees to which the consequent reference value C_j is believed to be true. If

$$\sum_{j=1}^{N} \beta_{jk} = 1 \quad (2)$$

is the k^{th} rule is said to be complete. Otherwise, it is incomplete. T_k is the total number of antecedent attributes used in k^{th} rule L is the number of all belief rules in the rule base. N is the number of all possible consequent in the rule base. For example a belief rule to assess Metabolic syndrome (A9) for PCOS can be written in the following way.

R_k: A2 is H^A3 is H^A4 is H^A5 is H THEN A9 is {((H,1.0),(M,0.0),(L,0.0)}

Where {(H (High), 1.0), (M (Medium), 0.0), (L (Low), 0.0)} is a belief distribution for A9 (masculinizing hormons) consequent, stating that the degree of belief associated with High is 100%, 0% with medium and 0% with low. In this belief rule, the total degree of belief is (1.0+0.0+0.0) =1, hence that the assessment is complete.

B. BRB Inference using ER

The ER approach [7] [18] [24]developed to handle multiple attribute decision analysis (MADA) problem having both qualitative and quantitative attributes. 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 consist 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 [15][16][18][24].

The input transformation of a value of an antecedent attribute 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 [14][24]. The ith value of an antecedent attribute 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 .The input value of P_j, which is the ith antecedent attribute of a rule, along with its belief degree ε_i is shown below by Eq. (3). The belief degree ε_i to the input value is assigned by the expert in this research.

$$H(P_i, \varepsilon_i) = \{(A_{ij}, \alpha_{ij}), j = 1, \ldots, j_i\}, i = 1, \ldots\ldots, T_k \quad (3)$$

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 A_{ij} (ith value) is the j^{th} referential value of the input P_i . α_{ij} is the belief degree to the referential value A_{ij} with $\alpha_{ij} \geq 0$. $\sum_{j=1}^{j_1} \alpha_{ij} \leq 1\ (i = 1, \ldots, T_k)$, and j_i is the number of the referential values.

For example, the input 0.92 for masculinizing hormones is equivalently transformed to{(High,0.87), (Medium,0.11), (Low, 0.02)}.The input value of an antecedent attribute is collected from the expert in terms of linguistic values such as 'High', 'Medium' , and 'Low' .This linguistic value is then assigned degree of belief ε_i by taking account of expert judgment. This assigned degree of belief is then distributed in terms of belief degree α_{ij} of the different referential values A_{ij} [High, Medium, Low] of the antecedent attribute. The above procedure of input transformation is elaborated by equations (4 and 5) given below. However, it is important for us to know, with what degree of belief it is High and with what degree of belief it is Medium. This phenomenon can be calculated with the following Eq. (4) and Eq. (5).

$$\beta_{n,i} = \frac{h_{n+1} - h}{h_{n+1,i} - h_{n,i}}, \beta_{n+1,i} = 1 - \beta_{n,i} \quad (4)$$

$$If\ h_{n,i} \leq h \leq h_{n+1,i} \quad (5)$$

Here, the degree of belief $\beta_{n,i}$ is associated with the evaluation grade Low while$\beta_{n+1,i}$ is associated with the upper level evaluation grade i.e. High.

When the k^{th} rule is activated, the weight of activation of the k^{th} rule, w_k is calculated by using the flowing formula [17][18].

$$w_k = \frac{\theta_k \alpha_k}{\sum_{j=1}^{L} \theta_j \alpha_j} = \frac{\theta_k \prod_{i=1}^{T_k} (\alpha_i^k)^{\bar{\delta}_{ki}}}{\sum_{j=1}^{L} \theta_j [\prod_{l=1}^{T_k} (\alpha_l^k)^{\bar{\delta}_{kl}}]} \quad (6)$$

here,

$$\bar{\delta}_{kl} = \frac{\delta_{ki}}{max_{i=1,\ldots,T_k}(\delta_{ki})}$$

Where $\bar{\delta}_{kl}$ is the relative weight of P_i used in the k^{th} rule, which is calculated by dividing weight of P_i with maximum weight of all the antecedent attributes of the k^{th} rule. By doing so, the value of $\bar{\delta}_{kl}$ becomes normalize, meaning that the range of its value should be between 0 and 1. $\alpha_k = \prod_{i=1}^{T_k} (\alpha_i^k)^{\overline{\delta_{ki}}}$ Is the combined matching degree, which is calculated by using the multiplicative aggregation function.

When the k^{th} rule as given in Eq. (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 $\bar{\beta}_{ik}$ in the i^{th} consequent C_i of the k^{th} rule is updated based on the actual input information as [15] [17] [18][24].

$$\beta_{ik} = \bar{\beta}_{ik} \frac{\sum_{t=1}^{T_k} (\tau(t,k) \sum_{j=1}^{j_t} \alpha_{tj})}{\sum_{t=1}^{T_k} (\tau(t,k)} \quad (7)$$

where,

$$(t,k) = \begin{cases} 1, if\ P_i\ is\ used\ in\ defining\ R_k (t = 1, \ldots \ldots, T_k) \\ 0, otherwise \end{cases}$$

Here, $\bar{\beta}_{ik}$ is the original belief degree and β_{ik} is the updated belief degree.

Due to the incomplete input for "masculinizing hormons",the belief degree of the connected rules needs to be modified to show the incompleteness by using Eq. (7).

$$\beta_{ik} \equiv \bar{\beta}_{ik} \frac{1.6}{2} = \bar{\beta}_{ik} * 0.8, \ i = 1,2,3; k = 1, \ldots \ldots, 9$$

Therefore $0 < \sum_{i=1}^{3} \beta_{ik} < 1$ for all rules that are associated with "masculinizing hormons". Using the sub rule base, the assessment result for "masculinizing hormons" is obtained using BRBES system :{(High, 0.66), (Medium ,0.23), (Low, 0.02), (Unknown, 0.09)} where Unknown in the above result means that the output is also incomplete input. ER approach is used to aggregate all the packet antecedents of the L rules to obtain the degree of belief of each referential values of the consequent attribute by taking account of given input values P_i of antecedent attributes. This aggregation can be carried out either using recursive or analytical approach. In this research analytical approach [19] [24] has been considered since it is computationally efficient than recursive approach [14][20] [21][24], because analytical approach deal with all parameter such as rule weight, attribute weight, belief degree, utility etc. For this why there is no chance of absence of any parameter. The conclusion *O(Y),* consisting of referential values of the consequent attribute, is generated. Eq. (8) as given below illustrates the above phenomenon.

$$O(Y) = S(P_i) = \{(C_j, \beta_j), j = 1,2, \ldots \ldots, N\} \quad (8)$$

Where, $\beta_j$denotes the belief degree associated with one of the consequent reference values such as C_j and $\beta_j$is calculating by analytical format of the ER algorithm [3] [24]as illustrated in equation (9).

$$\beta_j = \frac{\mu[\prod_{k=1}^{L}((w_k \beta_{jk} + 1 - w_k \sum_{j=1}^{N} \beta_{jk})) - \prod_{k=1}^{L}(1 - w_k \sum_{j=1}^{N} \beta_{jk})]}{1 - \mu[\prod_{k=1}^{L} 1 - w_k]} \quad (9)$$

With

$$\mu = \left[\sum_{j=1}^{N} \prod_{k=1}^{L} \left(w_k \beta_{jk} + 1 - w_k \sum_{j=1}^{N} \beta_{jk}\right) - (N-1) \prod_{k=1}^{n} (1 - w_k \sum_{j=1}^{N} \beta_{jk})\right]^{-1}$$

The final combined result or output generated by ER is represented by $\{(C_1, \beta_1), (C_2, \beta_2), \ldots \ldots, (C_N, \beta_N)\}$

Here β_j is the final belief degree attached to the j^{th} referential value C_j of the consequent attribute, obtained after combining all activated rules in the BRB by using ER.

C. Output of the BRB System

The output of the BRB system is not crisp/numerical value. Hence, this output can be converted into crisp/numerical value by assigning a utility score to each referential value of the consequent attribute [17][24].

$$H(A^*) = \sum_{j=1}^{N} u(C_j) B_j \quad (10)$$

where, $H(A^*)$ is the expected score expressed as numerical value and $u(C_j)$ is the utility score of each referential value. For example, in this paper the overall assessment result is {(H, 0.55), (M, 0.25), (L, 0.20)} for PCOS disease, then the expected utility score is 0.675 or 68% which represents Medium risk disease. In this paper the RIMER methodology to address various type of uncertainty such as incompleteness, ignorance and impreciseness by using Eq. (7) and Eq. (11).The incompleteness as mentioned occurs due to ignorance, meaning that belief degree has not been assigned to any specific evaluation grade and this can be represented using the equation as given below.

$$\beta_H = 1 - \sum_{n=1}^{N} \beta_n \quad (11)$$

where, β_H is the belief degree unassigned to any specific grade. If the value of β_H is zero then it can argued that there is an absence of ignorance or incompleteness. If the value of β_H is greater than zero then it can be inferred that there exists ignorance or incompleteness in the assessment.

$$\beta_{ik} = \bar{\beta}_{ik} \frac{\sum_{t=1}^{T_k} (\tau(t,k) \sum_{j=1}^{j_t} \alpha_{tj})}{\sum_{t=1}^{T_k} (\tau(t,k)}$$

III. BRBES ARCHITECTURE

Architectural design represents the structure of data and program components that are required to build a computer-based system. It is also considers the pattern of the system organization, known as architectural style. BRBES adopts the three-layer architecture, which consisting of presentation (user interfaces), application (Inference engine) and data processing layer (Belief rule Base) is as shown in Fig. 1.

A. System Components

The input clarifications of input antecedent are

A1=menstrual disorders,
A2= acne
A3=hirsutism
A4= hyperorrheamen
A5= androgenic alopecia
A6= Central obesity
A7= Insulin resistance are transformed to referential

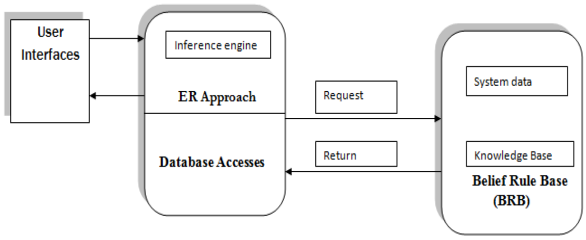


Fig. 1. BRBES Architecture.

Value is evaluated by Eq. (4), (5) on behalf of expert. The input clarifications of this BRB system transformed to referential is shown in Table I.

TABLE I. THE INPUT ARE TRANSFORMED IN REFERENTIAL VALUE

Sl. No.	Input Antecedent	Expert Belief	Referential Value		
			High	Medium	Low
0	A1	1.0	1	0	0
1	A2	0.5	0.1	0.7	0.2
2	A3	0.8	0.5	0.5	0
3	A4	0.5	0.1	0.8	0.1
4	A5	1	0.8	0.2	0
5	A6	0.5	0.1	0.4	0.5
6	A7	1	0.8	0.2	0

B. Knowledge Base Constructed using BRB

In present paper, we worked on polycystic ovary syndrome (PCOS). In order to construct BRB knowledge base of this system we designed a BRB framework to PCOS diagnosis according to clinical domain expert (doctor).The BRB framework of PCOS diagnosis as illustrated in Figure 2, From the framework, it can be observed that input factors that determine this disease level. The BRB knowledge base has different traditional rule to assessment, which need to convert brief rules.

In such situations, belief rules may provide an alternative solution to accommodate different types and degrees of uncertainty in representing domain knowledge. A BRB can be established in the following four ways[15]- (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 constructed initial BRB by the domain expert knowledge. There will be four sub-rule bases, which can be named as follows:

1) A8 sub-rule-base
2) A9 sub-rule-base
3) A10 sub-rule-base
4) A11 sub-rule-base

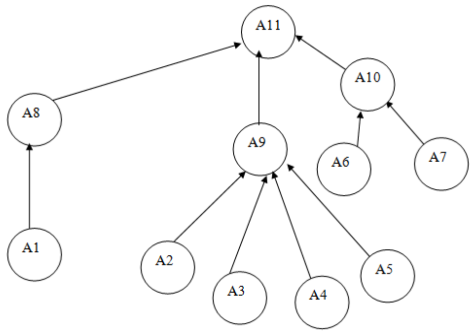


Fig. 2. Hierarchical Relationship among PCOS diagnosis variables.

The Calculation of the initial rule-base for various sub-rule-bases

Initial rule-base for A8 which will be consisting of 3 rules
Initial rule-base for A9 which will be consisting of 81 rules
Initial rule-base for A10 which will be consisting of 9 rules
Initial rule-base for A11 which will be consisting of 27 rules

The entire BRB consists of (3+81+9+27) =120 belief rules. It is assumed that all belief rules have equal rule weight; all antecedent equal weight, and the initial belief degree assigned to each possible consequent by two expert from accumulated the data. To better handle uncertainties, each belief rule considered the three referential values are High (H), Medium (M) and Low (L)

TABLE II: INITIAL BELIEF RULES OF SUB-RULE-BASE (METABOLIC SYNDROME)

Rule No.	Rule Weight	IF		THEN		
		A6	A7	A10(Metabolic syndrome)		
				High	Medium	Low
0	1	H	H	0.8	0.2	0
1	1	H	M	0.4667	0.5333	0
2	1	H	L	0.0667	0.9333	0
3	1	M	H	0	0.9333	0.1
4	1	M	M	0	0.8	0.2
5	1	M	L	0	0.6667	0.3
6	1	L	H	0.3333	0.6667	0
7	1	L	M	0	0.9333	0.1
8	1	L	M	0	0.8	0.2

C. *Inference Engine using ER*

This BRBES designed using the ER approach [15] [18] [20] [21] which is described in section II (B). It is similar to traditional forward chaining. The inference with a BRB using the ER approach also involves assigning values to attributes, evaluating conditions and checking to see if all of the conditions in a rule are satisfied. The BRB inference process using the ER approach described by the following steps are input transformation, calculation of the activation weight, calculating combined belief degrees to all consequents, belief degree update and aggregate multiple activated belief rules. The inputs of data are of two types, objective and subjective. Input transformation of this system and input clarification are deduced in previous section and Table I by using Eq. (4) and Eq. (5). After the value assignment for antecedent, calculating the combined matching degrees between the inputs and the rule's antecedents, the next step is to calculate activation weight for each packet antecedent in the rule base using Eq. (6). The belief degrees in the possible consequent of the activated rules in the rule base are updated using Eq. (7). Then aggregating all activated rules using the ER approach to generate a combined belief degree in possible consequents using Eq. (8) and Eq. (9). Then expected result of PCOS diagnosis was calculated from its different consequents of factors. Finally, presenting the system inference results of PCOS diagnosis consequent which is not crisp/numerical value, then it is converted into crisp/numerical value for recommendation using Eq. (10).

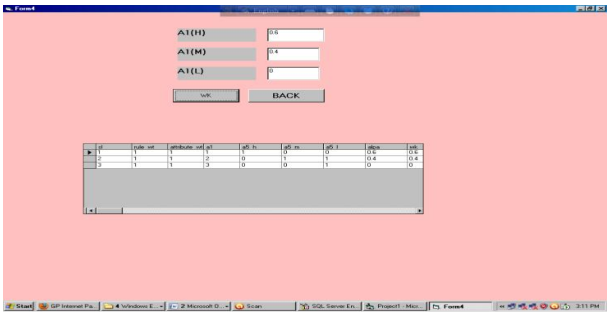


Fig. 3. Interface of the BRBES.

D. *BRB ES Interface*

System interface is an intermediate position that represents the interaction between user and system. Fig. 3 represents the BRB system interface of this paper.

IV. RESULTS AND DISCUSSION

The clinically simulated data set variables used in this paper to determine PCOS are one patient's clinical signs, symptoms and PCOS include namely menstrual disorders, acne, hirsutism, hyperorrheamen, androgenic alopecia, Central obesity, Insulin resistance, A11(PCOS) A8 (infertality), (A9) masculinizing hormons, (A10) Metabolic syndrome. Where the PCOS is dependent variable, it is used to present outcome as having PCOS present or not, if PCOS diagnosis score result is greater than 70% (High), if score result is less than 70% (Medium) otherwise Low. The real data set of 36 patients were collected and simulated having PCOS with different clinical level. The eight patient's simulated data set with diagnosis outcome is presented as example in Fig. 4. This figure represents overall PCOS diagnosis outcome from patient's information. The result of this system is measured in percentage for recommendation. The output of this system was generated based on output utility Eq. (10). In this paper, the utility score of 100% assigned to 'High', 50% assigned to 'Medium', and 0% assigned to 'Low'. For example, we can estimate overall system output PCOS as 99%, if the Fuzzy result of the system is {(High, 0.90), (Medium, 0.10), (Low, 0)}.

In the case study, the PCOS of 36 patients using this system, doctors' manual system and clinical history result is shown in Fig. 4. The clinical historical results were considered as benchmark. From Fig. 4 it can be observed that ES generated result has less deviation than from benchmark result. Hence, it can be argued that ES output is more reliable than manual system. Therefore, it can be concluded that if the assessment of PCOS is carried out by using the ES, eventually this will play an important role in taking decision to avoid unnecessary costly lab investigation.

Patient ID	Age	A1	A2	A3	A4	A5	A6	A7	Manuel Result	Benchmark Result	BRB System Result	Risk Stage
1	50	H	H	H	H	H	M	L	80.44%	92.55%	91.55%	High>70%
2	67	M	L	M	H	M	L	M	67.33%	69.55%	66.99%	Medium<70%
3	47	L	H	M	L	M	H	H	57.55%	50%	52.66%	Low<50%
4	52	M	M	H	L	H	M	L	85%	90%	87%	High>70%
5	40	L	L	L	M	L	L	H	52.11%	55.23%	54%	Low<50%
6	53	H	H	H	H	H	L	M	85%	90%	87%	High>70%
7	36	M	L	M	H	M	Y	L	52.11%	43.23%	45%	Low<50%
8	63	L	H	M	L	M	L	M	61.66%	67.43%	69.88%	Medium<70%

Fig. 4. Simulated Data by BRBES (H-High, M-Medium, L-Low).

V. CONCLUSION

The development and application of a belief rule based Expert system (BRBES) to detect Polycystic Ovary Syndrome (PCOS) by using signs and symptoms of patients have been presented. The prototype ES is embedded with a novel methodology known as RIMER which allows the handling of various types of uncertainty. Hence this paper considered BRB as a robust tool for detecting PCOS. Consequently, the prototype ES can handle various types of uncertainties found in clinical domain knowledge as well as in signs and symptoms of a patient. 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. It can also provide a percentage of risk recommendation, which is more reliable and informative than from the traditional expert's opinion. The prototype ES can only be used to detect PCOS by using signs and symptoms of a patient.

In further, research will be considered the learning module system. The validation of the system is necessary by using real patient historical data and the knowledge representation parameters should be trained by real clinical data using the BRB optimal learning training model. At last, employing the RIMER to develop BRBES for PCOS suspicion which is a valid new approach and real data can be used to training in future research.

Acknowledgment

The authors are grateful to Dr. Mumtahena Mahmuda MBBS (DMC), FCPS, BCS (Health) working as the medical officer of government Hospital for providing domain knowledge and support in this research.

REFERENCES

[1] M Sonmez, G. Graham and J. B. Yang and G D Holt, "Applying evidential reasoning to pre-qualifying construction contractors", Journal of Management in Engineering, Vol.18, No.3, pp.111-119, 2002.

[2] J. B. Yang, "Rule and utility based evidential reasoning approach for multiple attribute decision analysis under uncertainty", European Journal of Operational Research, Vol. 131, No.1, pp.31-61, 2001.

[3] Y. M. Wang, J. B. Yang and D. L. Xu, "Environmental Impact Assessment Using the Evidential Reasoning Approach", European Journal of Operational Research, Vol.174, No.3, pp.1885-1913, 2006.

[4] Lisa M. (2008). The Sage encyclopedia of qualitative research methods. Los Angeles, Calif.: Sage Publications. ISBN 1-4129-4163-6.

[5] http://www.pearson.ch/1449/9780273722595/An-Introduction-to-Geographical.aspx.

[6] Dodge Y. (2003) The Oxford Dictionary of Statistical Terms, OUP. ISBN 0-19-920613-9.

[7] D. L. Xu and J. B. Yang, "Introduction to multi-criteria decision making and the evidential reasoning approach", Working Paper Series,Paper No.: 0106 , ISBN: 1 86115 111 X , Manchester School of Management, UMIST, pp. 1-21, 2001.

[8] A. Taroun and J. B. Yang. "Dempster-Shafer theory of evidence: potential usage for decision making and risk analysis in construction project management." Journal of the Built and Human Environment Review 4, no. 1(2011) : 155-166.

[9] L. Zadeh, A simple view of the Dempster-Shafer Theory of Evidence and its implication for the rule of combination, The Al Magazine, Vol. 7, No. 2, pp. 85-90, Summer 1986.

[10] Kari Sentz and Scott Ferson (2002); Combination of Evidence in Dempster–Shafer Theory, Sandia National Laboratories SAND 2002-0835.

[11] Bragge, J.; Korhonen, P., Wallenius, H. and Wallenius, J. (2010). "Bibliometric Analysis of Multiple Criteria Decision Making/Multiattribute Utility Theory". IXX International MCDM Conference Proceedings, (Eds.) M. Ehrgott, B. Naujoks, T. Stewart, and J. Wallenius,. Springer, Berlin 634: 259–268.

[12] P. Sen and J. B. Yang, "Design decision making based upon multiple attribute evaluation and minimal preference information", Mathematical and Computer Modeling, Vol.20, No.3, pp.107-124, 1994.

[13] B. Yang and P. Sen, "Multiple attribute design evaluation of large engineering products using the evidential reasoning approach", Journal of Engineering Design, Vol.8, No.3, pp.211-230, 1997.

[14] Mahmud, T , Hossain, M.S(2012)"An Evidential Reasoning-based Decision Support System to Support House Hunting", International Journal of Computer Applications 57(21):51-58, November 2012.Published by Foundation of Computer Science, New York, USA.

[15] Yang, J. B., Liu, J., Wang, J., Sii, H. S. & Wang,H.W.(2006) belief rule-base inference methodology using theevidential reasoning approach - RIMER. IEEE Transactions onSystems Man and Cybernetics Part A-Systems and Humans,36, 266-285

[16] D.L. Xu, J. Liu, J.B. Yang, G.P. Liu, J. Wang, I.Jenkinson, J. Ren, Inference and learning methodologyof belief-rule-based expert system for pipeline leak detection,Expert Systems with Applications 32 (2007) 103–113.

[17] B. Yang, J. Liu, D.L. Xu, J. Wang, H.W. Wang, Optimal learning method for training belief rule based systems, IEEE Transactions on Systems, Man,and Cybernetics (Part A) 37 (2007) 569–585.

[18] Patwary, M.J.A , Akter, S, Mahmud, T (2014) "An Expert System to Detect Uterine Cancer under Uncertainty" IOSR Journal of Computer Engineering (IOSR-JCE), e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 16, Issue 5, Ver. I (Sep – Oct. 2014), PP 36-47)

[19] A. M. Nonvich and T. B. Turksen, "A model for the measuremnt of membership and the consequnces of its empical implementation," Fuzz, Sers Syst., vol. 12, pp. 1-25, 1984.

[20] Mahmud,T.;Rahman,K.N.;Hossain,M.S.(2013)."Evaluation of Job Offers Using The Evidential Reasoning Approach".Global Journal of Computer Science and Technology. Volume13 Issue 2 Version 1.0 . Published by Global Journals Inc.(USA).

[21] Mahmud,T , Sikder,J (2013) "Intelligent Decision System for Evaluation of Job Offers" 1st National Conferenceon Intelligent Computing and Information Technology(NCICIT) , November 21, CUET, Chittagong-4349,Bangladesh.

[22] Ankita Thakur,1,2 Vijay Mishra,2 and Sunil K. Jain ,Feed Forward Artificial Neural Network: Tool for Early Detection of Ovarian Cancer, Jul 5, 2011.

[23] Yu-Wang Chen, Jian-Bo Yang, Dong-Ling Xu, Zhi-Jie Zhou, Da-Wei Tang, Inference analysis and adaptive training for belief rule based systems, Expert Systems with Applications 38; 12845–12860, 2011.

[24] Hossain,M.S.,Khalid,M.S.,Akter,S.,& Dey,S.(2014).A Belief Rule-Based Expert System to Diagnosis influenza.In Proceeding of the 9th international Forum on Strategic Technology.(pp.113-116).IEEE Professional Communication Society.10.1109/IFOST.2014.6991084.

Drowsiness Level Detection for the Protection from Accident of Intelligent Transportation System (ITS)

Md. Kamrul Hasan*, Shantanu Sen Gupta, S. M. Hasnat Ullah, and Mohiuddin Ahmad
Department of Electrical and Electronic Engineering (EEE)
Khulna University of Engineering & Technology (KUET)
Khulna-9203, Bangladesh
*E-mail: m.k.hasan@eee.kuet.ac.bd

***Abstract*— Road accidents are the most common phenomenon all over the world and also most dominated in developing countries which are primarily caused due to driver fatigue. The key objective of this research is to find out the drowsiness of the driver to estimate an arousal level (Unconsciousness) and to apply this level of unconsciousness to the ITS (Intelligent Transportation System) that can provide a signal to the drivers to be alert at their high arousal state. In this research, we have proposed and implemented an ITS which can resist the vehicle from dangerous or unwanted conditions in the road. The concepts of fatigue is also discussed and offered a swift on psychophysiological associations with driver fatigue. When the level of drowsiness of driver's EEG (Electroencephalogram) signal is at the state of unwanted level, this ITS automatically sends a signal to the control gear to stop the system to prevent traffic accidents. Here, eye blink sensor is used to estimate drowsiness level from the driver which provides less noise than direct measurement and analysis of EEG signal for drowsiness level detection.**

Keywords*— ITS (Intelligent Transportation System); Eye Blink Sensor; Driver Fatigue; Drowsiness Level; Driving Alert; Arduino.

I. Introduction

Driver fatigue which is mainly caused due to sleeping or drinking condition of the driver is most dominated factor for the increasing number of accidents on today's roads [1], [2]. Driver fatigue has been one of the chief sources of traffic misfortunes all over the world. As reported in [3], in United Kingdom (UK) it was estimated that more than 20% of traffic accidents have been resulted from driver fatigue, while in the US there are around 50% of fatigue related to fatal accidents. As a consequence, some of the developing countries like Bangladesh, this road accident is more severe than developed countries. Hence, researchers have activated to pay more attentions to design ITS (Intelligent Transportation System) to provide driving safety and offer the way to the diminution of road cracks.

There are four major categories of demeanor that normally appears to the driver-

1. Normal demeanor
2. Drunk demeanor
3. Fatigue demeanor
4. Reckless demeanor

The driver's demeanor is on the state of normal when the concentration of the driver is fixed on the driving and at this state driver is more relaxed. In this state the controlling power and controlling speed of the vehicle is highest and avoiding of sudden accident is also high [4].

In this state of driving, a driver is drunk when the he is intoxicated by alcohol and is characterized by a set of observable actions like sudden acceleration, driving without controlling the speed [5].

At the fatigue demeanor state, driver drive a car about 15-20 hours without sleeping which behaves exactly as a driver who has 0.05% intoxication of alcohol. Based on this argument, fatigue driving is defined as driving that exhibits the same characteristics as drunk driving, but there is no alcohol intoxication in the blood of the driver [6].

It is the condition of demeanor of the driver who drives at high speed, with a high degree of acceleration and makes other traffic participants at risk. There is no alcohol intoxication and the driver's eyes are open, but the following behaviors such as driving with sudden acceleration, not maintaining the proper lane position and not controlling the vehicle's speed are exhibited [7].

This vital fatigue of driver can be detected using three major techniques [8] which are-

1. Physiological dimensions
2. Visual signs
3. Driving recital

The first two categories investigate the activities of the driver directly, whereas the third category investigates activities of the driver indirectly. Physiological investigation includes the following types which require electrodes (Dry or Wet) to driver to measure features such as

1. Electroencephalography (EEG) measurement
2. Electrooculography (EOG) measurement
3. Electrocardiography (ECG) measurement

An EEG-based system designed and implemented by [9] was able to detect fatigue which have error rate of roughly 10%, but the authors claim to be very reliable. Similar EEG-based studies were also directed by [10]-[12]. The main drawback is recording of EEG signal from the brain is the major challenge since contact impedance at the contact surface is the major problem for EEG recording due to the cause of contacting gel for wet electrode is drying with the increase of time and dry electrode based EEG measurement is so complex and noisy. Hence this system is both intrusive and impractical. The second way which is visual signs from a driver's face can also serve as an indicator of fatigue but this is too much costly. The third category named as driving recital of fatigue detection techniques which is the techniques that investigate how the driver knobs the vehicle. Fatigue can be detected by variations in the steering wheel angle, vehicle lateral position and vehicle speed. Such kinds of fatigue is designed by the authors in [13], [14].

The development of the techniques of detecting or mitigating drowsiness at the wheel is a major contest in the field of accident evading systems analysis. This is due to the hazard that drowsiness offerings on the road, methods need to be developed for neutralizing its affects. The goal is that to design and implement a prototype drowsiness detection system and take necessary steps to prevent the road accident for the developing countries [15].

II. PROPOSED MODEL AND ALGORITHM

The complete research work is described in several block diagram as given below-

A. *The proposed Topology of Intelligent Transportation System (ITS)*

ARDUINO is used as a main processor of our proposed ITS system which takes input from the control switch and eye blink sensor and process these signal finally send alarm to the buzzer. Motor and LCD also connected to the ARDUINO for the proper operation of motor and display the state of ARDUINO respectively. The proposed topology of Intelligent Transportation System (ITS) is shown in the Fig. 1.

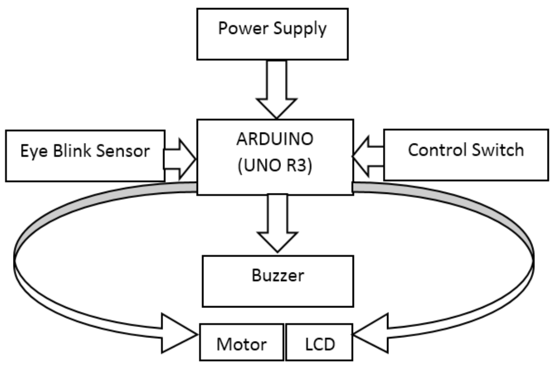


Fig. 1. Proposed topology of Intelligent Transportation System (ITS)

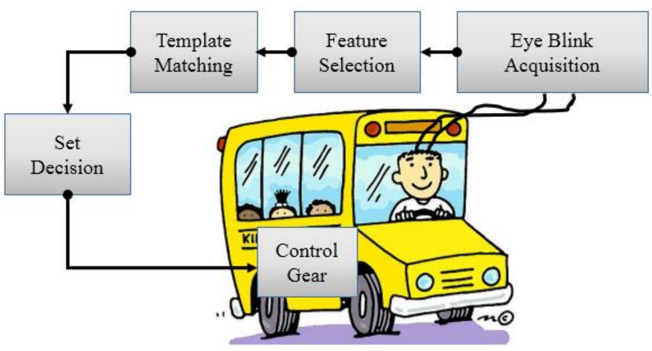


Fig. 2. Proposed block digram of ITS

B. *Intelligent Transportation System (ITS)*

ITS improve transportation safety and mobility which can also enhance the flexibility of safe journey. It mainly contains an Eye blinking sensor module. This module takes the data from driver's eye. Then ARDUINO processes these data and compares with desired data. After processing, controller sets decision that the vehicle should move or not. A typical block diagram of ITS is given in Fig. 2.

C. *Overall Flow diagram*

When driver starts his vehicle, the protection module turns on automatically. In day time, control switch should be pressed, then display shows 'Day mode' and vehicle can move. Otherwise display shows 'Welcome', eye blink sensor starts it's monitoring. If sensor output is less than desired output, it is considered as driver is conscious, display shows 'it's safe' and vehicle can move. Otherwise if sensor output is more than desired output, it is considered as driver is sleeping, display shows 'Dangerous', buzzer turns on and vehicle stops automatically. Then when driver awakes, he can take action by pressing control switch. An overall flow diagram is given in Fig. 3.

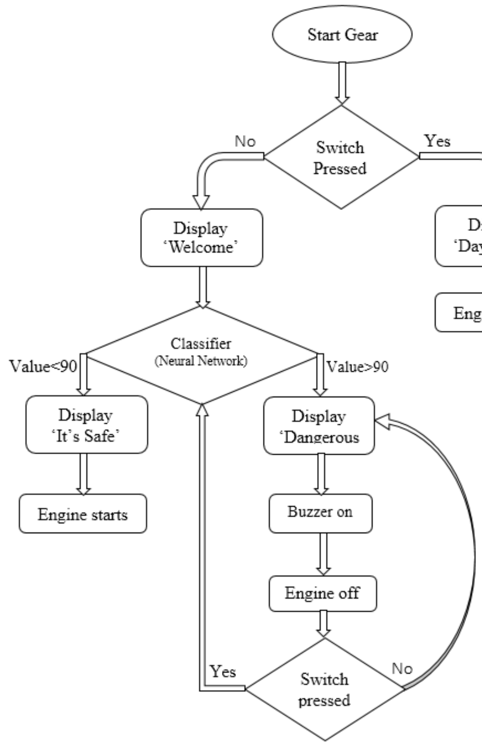


Fig. 3. Overall flow diagram for the proposed ITS

D. *Low cost eye blink signal measurement*

Eye blinking can be measured in various ways but the most cost effective way is using Eye blinking sensor made by Infrared Ray sensor pair. The eye blinking sensor module is attached with a glass by maintaining the angle with respect to eye. Then ARDUINO takes input signal from the output signal of sensor module. After that, the collected signal is compared with the standard signal to take decision either driver is sleeping or not. This is shown in Fig. 4.

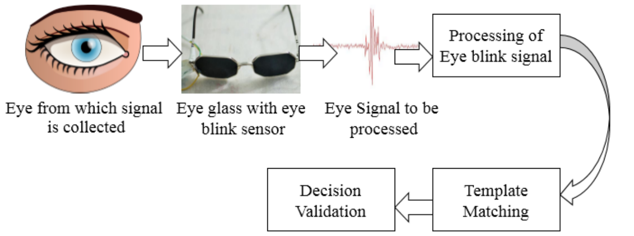


Fig. 4. Extarction of eye bilnk signal at drowsy condition

E. Algorithm for ARDUINO

The algorithm used for the ARDUINO, is described in several steps which is mention below step by step-

Step 1: Declare I/O pins in ARDUINO.

Step 2: If control switch pressed

Step 3: Display "Day mode".

Motor On.

Step 4: Else if sensor output< maximum open eye output

Step 5: Display "It's safe".

Motor On.

Step 6: Else if sensor output> maximum open eye output

Step 7: Display "Dangerous"

Buzzer on;

Motor off.

Step 8: If control switch pressed

Go to **Step 4**

Step 9: Else if control switch not pressed

Go to **Step 7**

F. Artificial Neural Network (ANN) Classifier

ANNs are processing devices or algorithms similar to the neurons of the brain of mammalian. For the applications in the classification ANN has greater priority due to their quick response. McCullogh-Pitts model of Artificial Neural Network (ANN) is shown in Fig. 5. The classifier inputs at the inputs terminal are added then passed through the thresholding function as shown in Fig. 6. In our research, ANN is used to take the classification of normal and abnormal state of driving condition and this classified decision is used to drive the car safely as followed in Fig. 3.

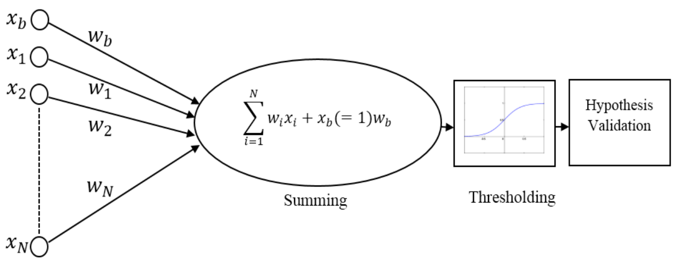


Fig. 5. McCullogh-Pitts model of Artificial Neural Network (ANN)

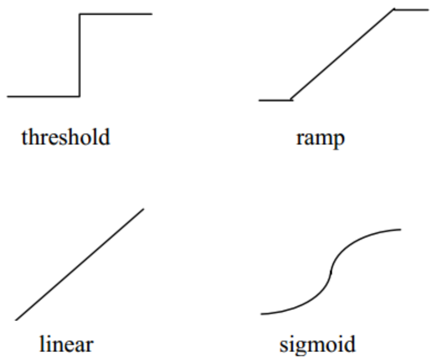


Fig. 6. Typically used thresholding function used in ANN

III. RESULTS AND DISCUSSION

The research result of this manuscript is described in several subsection which are described below-

A. Glass Design with eye blink sensor

The main part of the complete hardware setup is the glass design which contains eye blink sensor. Infrared Ray sensor pair is representing the eye blink sensor. This eye blink sensor has been attached with the glass. It can be made by perfect positioning from different angle. Here as a homemade, it has been attached under the frame maintaining almost 45 degree angle vertically such that the transmitting ray can reach the inside of eye.

Fig. 7. Eye blink sensor design with correct angle (left) and placement of sensor to driver (right)

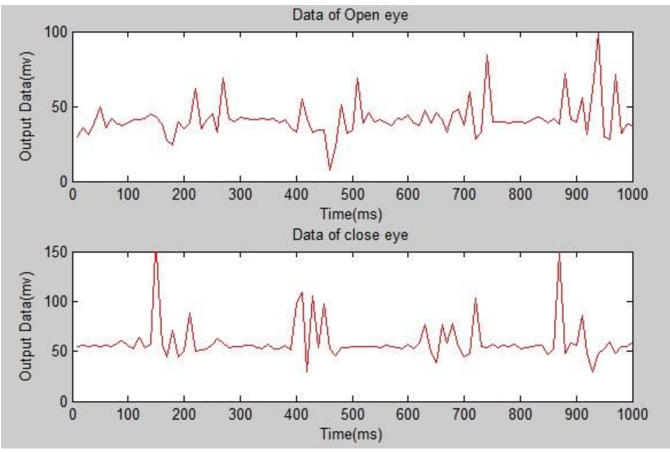


Fig. 8. Typical eye blink signal extracted from the eye blink sensor for eye closed state (down) and eye open state (up)

A typical Glass Design with eye blink sensor which is designed in our research are shown in the Fig. 7. Perfect positioning might give us more standard output. By the way, though there was the limitation of homemade prototype, a desired output has been gained. By processing and verifying this output, it has been decided that the vehicle should move or not. The extracted eye blink signal used for the controlling is shown in Fig. 8. There are two condition of extraction of eye blink signal one is at the eye closed state at Fig. 8 (Down) another is eye open state at Fig. 8 (Up).

B. Algorithm load into ARDUINO

ARDUINO is the controlling or processing unit of this whole project. Here ARDUINO Uno R3 has been used. Programming was not so quite easier, liquid crystal library was used for LCD. Overall programming has been done and loaded into ARDUINO using ARDUINO software version 1.6.0. A typical representation of ARDUINO is given in the Fig. 9.

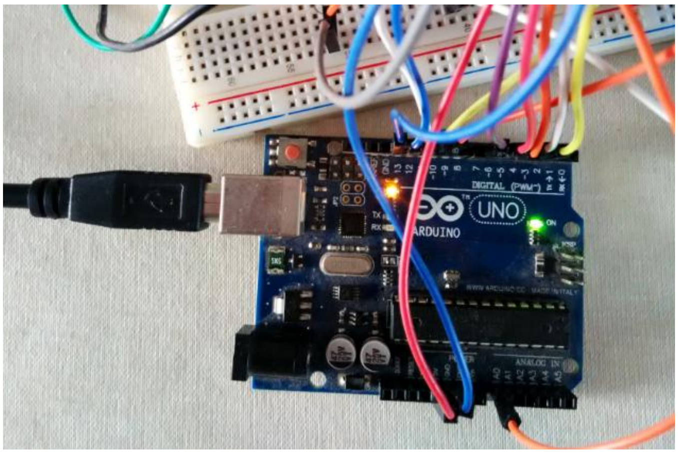


Fig. 9. ARDUINO to process and control the gear of ITS

C. Motor controller with L293D driver

A DC gear motor as shown in Fig. 10 was used to represent the sample of vehicle. L293D driver IC is needed because microprocessors works with low current but motor needs high current. So, this driver IC has been used. The controlling signal was received from ARDUINO.

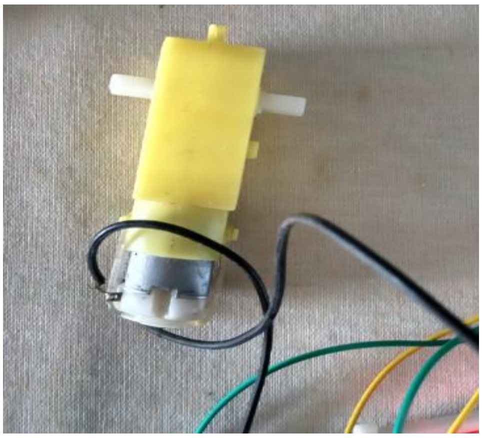


Fig. 10. Motor which is to be controlled

D. Overall hardware implementaion

First of all, eye blink sensor module has been designed. Then loading the program to ARDUINO it has been tested and collected the data. From different angle and in different condition it has been tested to set a standard value or reference value with which can be compared. Sensor output=90 was decided as our standard value. When it was responding the motor has been connected. In overall hardware, there also Buzzer and LCD have been used. Alarm has been listened from the buzzer when the vehicle was in dangerous situation. The condition of the vehicle was displayed by the LCD. The overall system is represented pictorially as Fig. 11.

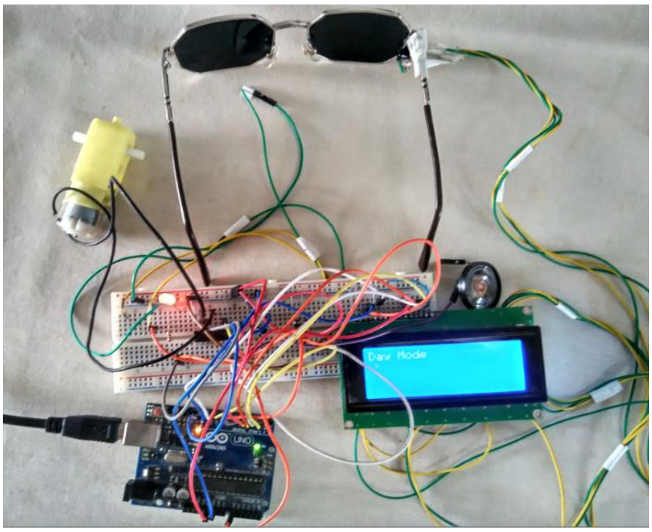


Fig. 11. Overall hardware implementation of proposed ITS

IV. CONCLUSSION

In our research manuscript, we have presented an Intelligent Transport System (ITS) which depends on the fatigue detection of the driver. Our design system is used to evade numerous road accidents that primarily caused due to the drowsiness of the driver all over the world especially in developing countries. Detecting and monitoring of the demeanor of driver is vital to provide safety against the road accident and alerting the driver. In this paper, a noninvasive technique has been designed which is user friendly and no training required for the subject to use this device. Eye blink sensor is the vital part of this device which extract actual drowsy signal from the eye and sends this signal to the ARDUINO for further processing. Practically eye blink sensor can be fixed in vehicle in front of driver, so that it can monitor that whether driver is unconscious or not. One of the vital future application of our research is, this ITS system may also imply to the pilot of the aircraft to prevent aircraft accident. The performance of our designed ITS can be enhanced by using sophisticated eye blink signal processor which helps to find out the correct features at the drowsy state of the driver.

REFERENCES

[1] Z. Zhang, and J. S. Zhang, “Driver Fatigue Detection Based Intelligent Vehicle Control”, 18th International Conference on Pattern Recognition (ICPR), Vol. 2, pp. 1262-1265, Hong Kong, 2006.
DOI: 10.1109/ICPR.2006.462

[2] M. Eriksson, and N. P. Papanikolopoulos, “Eye-Tracking for Detection of Driver Fatigue”, Proceeding of the International Conference on Intelligent Transportation System, Boston, MA, pp.314-319, November 1997.

DOI: 10.1109/ITSC.1997.660494

[3] Smart Motorist, Inc., “Driver Fatigue is an Important Cause of Road Crashes”http://www.smartmotorist.com/traffic-and-safety-guideline/driver-fatigue -is- a-important-cause-ofroadcrashes.html (visited, 2011/08/22)

[4] L. M. Bergasa, J. Nuevo, M. A. Sotalo, and M. Vazquez, “Real-time system for monitoring driver vigilance”, IEEE Transactions on Intelligent Transportation Systems, Vol. 7, no. 1, pp. 66- 77, March 2006.

DOI: 10.1109/TITS.2006.869598

[5] E. Rogado, J. Garcia, R. Barea, L. Bergasa, and E. Lopez, “Driver Fatigue Detection System”, Proc. IEEE Int. Conf. Robotics and Biomimetics, Bangkok, Thailand, pp. 1105 – 1110, 22-25 February, 2009.

DOI: 10.1109/ROBIO.2009.4913155

[6] Bouchner, R. Pieknik, S. Novontny, J. Pekny, and M. Hajny, C. Borzová, “Fatigue of car drivers - detection and classification based on experiments on car simulators,” Proc. 6th Int. Conf. Simul., Model., Optim., Lisbon, Portugal, pp. 727–732, September 2006.

[7] D. Sowmya, I. Suneetha, and N. Pushpalatha, “Driver Behavior Monitoring through Sensors and Tracking the Accident using Wireless Technology”, International Journal of Computer Applications (0975 – 8887), Vol. 102, No. 2, September 2014.

[8] M. S. Devi, and P. R. Bajaj, "Driver Fatigue Detection Based on Eye Tracking", 1st International Conference on Emerging Trends in Engineering and Technology (ICETET), Nagpur, Maharashtra, pp. 649-652, July 2008. DOI: 10.1109/ICETET.2008.17

[9] N. L. Haworth, T. J.Triggs, and E. M Grey, “Driver Fatigue: Concepts, Measurement and Crush Countermeasures”, Human Factors group, Department of Psychology, Monash University, 1988.

[10] B. Jap, S.Lal, P. fischer, and E. Bekiaris, “Using EEG Spectral Components to Access Algorithms for Detecting Fatigue”, Expert Systems with Applications, vol. 36, no. 2, pp. 2352-2359, March 2009.

[11] A. Vuckovic, V. Radivojevic, A. Chen, and D. Popovic, “Automatic Recognition of Alertness and Drowsiness from EEG by an Artificial Neural Network”, Medical Engineering and Physics, vol. 24, pp. 349-360, March 2002.

[12] K. Shen, X. Li, C. Ong, S. Shao, and W. S. E, ”EEG Based Mental Fatigue Measurement Using Multiclass Support Vector Machines with Confidence Estimate”, Clinical Neurophysiology, vol. 119, pp. 1524-1533, May 2008.

[13] R. Sayed, and A. Eskandarian, “Unobtrusive drowsiness detection by neural network learning of driver steering”, Proceedings of the Instituion of Mechanical Engineers, Part D: Journal of Automobile Engineering, vol. 215, no. 9, pp. 969–975, June 2001.

[14] A. Giusti, C. Zocchi, and A. Rovetta, “A noninvasive system for evaluating driver vigilance level examining both physiological and mechanical data”, IEEE Transactions on Intelligent Transportation Systems, Vol. 10, no. 1, pp. 127-134, March 2009.

[15] K. Srijayathi, and M. Vedachary, “Implementation of Driver Drowsiness Detection System”, International Journal of Science, Engineering and Technology Research (IJSETR), Vol. 2, no. 9, September 2013.

Statistical Prediction Model of Rain and Dust Storm Worst Month in Microwave - Millimeter Wave Band

Mohd. Ghazali Hamza
Razak School of Engineering and Advanced Technology
Universiti Teknologi Malaysia
Kuala Lumpur, Malaysia
ehkmohamed2@live.utm.my

Liza.A.Latiff
Razak School of Engineering and Advanced Technology
Universiti Teknologi Malaysia
Kuala Lumpur, Malaysia

Yusaini H.Mohamed
Malaysia-Japan International Institute Technology
Universiti Teknologi Malaysia
Kuala Lumpur, Malaysia

Yusaini H. Mohamed , M.F. Camara
Faculty of Mechanical Engineering
Universiti Teknologi Malaysia
Kuala Lumpur, Malaysia

***Abstract*: - Recent development in the field of advanced communication services and applications has triggered interest in super high frequencies and significantly increased research activity in millimeter wave. In this paper, a new method of calculation of worst month and statistical model has been developed for use in the rain and dust storm to statistically predict the probability of exceedance are presented. A thorough review of various prediction models previously developed for rain or sand and dust storm are comprehensively carried out. Traditionally, the meteorological conditions are difficult to predict due to uncontrolled variations. Statistical distributions have been fit-to-data type with little or no physical insight. However, studies in the statistical modelling of climate data rain and sand and dust storm for attenuation signal in literature are limited. The new model was fed with data collected from Sudan in five main climatic zones at specific areas between 23°.10″.0′-37°.15″.0′ E and 21°.02″.0′ - 09°.05″.0′ N. Moreover, formula for prediction of the worst- month long-term statistics for different weather induced propagation impairments was developed. The knowledge of weather phenomenon is usually available several periods in advance, which has the potential to improve or eliminate losses due to optimized convergence in tropical, arid, and desert. The main contribution of this study is to assist in the design of both terrestrial and space radio links using frequencies above 10 GHz in range of microwave and millimeter wave bands.**

Keywords: Worst- Month Propagation, Probability, Attenuation, Sand and Dust storm; Microwave & Millimeter-wave.

I. INTRODUCTION (*HEADING 1*)

Radio links operating at frequencies above 10 GHz are subject to severe propagation impairment. Over the few earlier decades, radio communication services and applications underwent broad development. The demands that a radio spectrum has to satisfy are larger through the day. Having in intellect just right quality and fee powerful solutions, a brand new radio sys-tem has to be designed cautiously from the very commencing [1-4]. One of the most important characteristics of the propagation of signal are environment of the path (attenuation) loss and an accurate prediction of the attenuation caused by rain or dust storm are essential for the design of communication systems. Moreover, an accurate estimation of the propagation losses provides a good basis for a proper selection of base station locations and a proper determination of the frequency plan. In recent years have become parts and different regions of the world are suffering from climate change, which in turn became the affect of service operations, which must be studied and know the deal, including phenomena such as ice in places were not out, as well as dust storms[5-6]

Rainfall and dust storm prediction is one of the most imperatives, important and demanding operational tasks and challenge made by meteorological services around the world. It is important to consider many technical issues before going to design and establish an expensive Wireless system. It is necessary to see the mathematical predictions or calculations of different parameters before going to design such type of systems. The quick growth in Links in wireless communications made it possible for the different area around the world as well as rural people to access the internet and cellular phones as well other applications. How-ever, this growth came with many inherent problems. Some of which include poor signal transmission or reception in various climatic conditions.

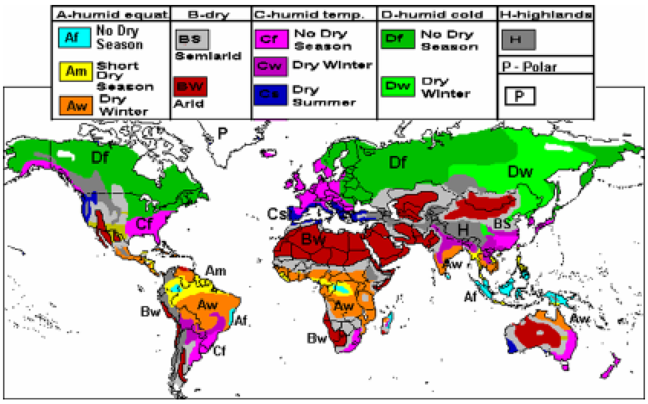


Figure 1 world's different climates zones tropical and arid.

Microwave and Millimeter wave bands links in space and terrestrial are prone to variations in attenuation over a large range due to rain and dust storm in links located in a raid and a tropical region, in some bad cases become more than 3 dB/km. traditionally, it has been considered to be part of meteorological conditions, unavoidable variations. Propagation conditions vary from month to month and from year to year, and the probability of occurrence of these conditions may vary by as much as several orders of magnitude. Some phenomena occur only rarely, requiring many years of observation to make any conclusions. For instance, elevated ducting may occur only several times per year in some locations, and in many areas, rain intense enough to affect propagation paths occurs for less than 1 percent of a year.

The outstanding sections of this paper are as follows: Effect of SDS to the higher frequency band and Microwave links is described in Section II. In Section III, we present the attenuation and development of prediction models and discussions. Finally, we end this study in Section IV with conclusion and recommendation and future work.

II. Relevant Studies

Relevant Studies: Rain attenuation is the dominant propagation impairment at frequencies above about 10 GHz. In addition, other impairments such as gaseous absorption, cloud attenuation, melting layer attenuation, and a tropospheric refractive effect becomes increasingly important with increasing operating frequency. A number of prediction models are available for the estimation of individual components [1]-[4]. However, methodologies that attempt to combine them in a cohesive manner are not widely available. This is partly due to the paucity of reliably measured data required to compare and verify such approaches.

Estimates statistics of various propagation effects that should be considered in the design of earth-space links. Propagation effects such as absorption, scattering and depolarization by hydrometeors, absorption due to atmospheric gases, multipath effects and ionosphere effects (typically only notable below 1GHz) can cause signal fade and need to be considered when implementing a satellite system to maintain a quality of service. Statistics for propagation effects provide an attenuation cumulative distribution function (CDF), which can be combined with further ITU-R recommendations to create an overall average annual attenuation CDF. Other ITU-R recommendations include rainfall rate, P.837-5, rain attenuation, P.838-3, cloud attenuation, P.840, and gas attenuation, P.676. Significant of study it is presence of the various forms of precipitation such as rain, snow, cloud and fog in a radio wave or microwave path are always capable of producing major impairments to terrestrial communications. It is, therefore, necessary to identify and predict the overall impact of every significant attenuation effect along any given path. A procedure for predicting the combined effect of meteorological factors attenuation and several other propagation impairments along earth or space links is presented.

There are several types of weather forecasts made in relation to time:

- A short-range forecast is a weather forecast made for a time period up to 48 hours.
- Medium range forecasts are for a period extending from about three days to seven days in advance.
- Long-range forecasts are for a period greater than seven days in advance but there are no absolute limits to the period.

Table: Summary of state

Years	References	Locations	Highlights
2000-2010	Bashir[1] Yagasena[1] Emiliani[2] Owolawi[3] Villiers[4] CHEN Crane[8]	Malaysia(1) Colombia(1) South Africa(1) Abu Dhabi(1) U.S.A(1) Taiwan (1) Sudan(1)	• To study monthly and seasonal rainfall and visibility, worst month and average worst month statistics of rain, and dust storm for the site under study. • worst month distribution estimation problem because distributions may be predicted for each of the months in a year and the worst one is then readily apparent
2010 and after	Musa [5]Ting[9] Thorvaldsen[6]	Norway(1) Malaysia(1) Nigeria(1) Iran(1) Iraq(1) Syria(1)	• The worst month concept is relevant to the higher grade telecommunications services under which propagation engineers must design radio systems which ultimately determine link availability.

Meteorological data are used to give an indication of the statistical rate of dust storm occurrence of a given of cross talk. As we mention in literature dusty seriously influences the performance of a communications satellite and ground link. The concept of worst-month plays an important role in space and earth links design where there is a need to know the design mar-gin that must be met in any particular month of the year. Worst-month statistics can be applied to quantities such as dust storm attenuation, visibility and polarization planes. The worst-month statistics for a given link are obtained by compiling a composite curve using the highest exceedance probability obtained in any calendar month at each threshold level.

The worst-month and annual statistics is related by the following ratio:

$$Q = X/Y \quad (1)$$

Where X is the average worst-month probability and Y is the average annual probability for the same threshold. Q is a function of the occurrence month level and the climatic region. Similar climatic regions will have similar values of Q in percentage. From a long term (1990- 2014) dust storm data more than twenty-five years data , we could find the average of the highest frequency of occurrence of dust storm events happens in the June in Capital Khartoum. Hence, June is the worst–time statistics. Optical visibility is directly related to the se-verity of dust-storm. Low visibility implies high number concentration of particles while low number concentration of particles represents high visibility the following model [1]-[8-9] gives the monthly probability of visibility to be less than a distance (li):

$$P\,(V<li) = [(ST\times t\times li)] / (30\times 24\times L)] \times 100\% \quad (2)$$

Where, P is the probability of exceedance in %, t is average dust-storm duration in hours, V is optical visibility in meters, T is an average number of storms per worst-month, and L is the maximum length of visibility considered in meters. We have been able to trace from data obtained from the Sudan Meteorological Authority (SMA -2014) that the visibility ranges from 0.1 km to 5 km and more in very few cases. The maximum length of visibility (L) considered in this case 1000 m. The table below depicts the visibility and its corresponding probability statistic results of the worst-month.

Worst-Month Visibility Statistics

For a preselected threshold level, the worst-month is defined as the month with the highest probability of exceeding the threshold level. Optical visibility is directly related to the severity of dust-storm. Low visibility implies high number concentration of particles while low number concentration of particles represents high visibility.

III. Methods

The methodology of your paper should be described in this section. The fading and enhancement distributions for the average worst month obtained from the methods can be converted to distributions for the average year by employing the following procedure: Step 1: Calculate the percentage of time PW fade depth A is exceeded in the large tail of the distribution for the average worst month as appropriate. Step 2: Calculate the logarithmic geo-climatic conversion factor □G from:

$$\Delta G = 10.5 - 5.6\log\left(1.1 \pm |\cos 2\xi|^{0.7}\right) - 2.7\log d + 1.7\log\left(1+|\varepsilon_p|\right) \quad \text{dB} \quad (3)$$

where $\Delta G \leq 10.8$ dB and the positive sign is employed for $\xi \leq 45°$ and the negative sign for $\xi > 45°$ and where: ξ: latitude (°N or °E) and d:path length (km) $|\varepsilon_p|$: the magnitude of path inclination. Calculate the percentage of time p fade depth A is exceeded in the large fade depth tail of the distribution for the average year from:

$$p = 10^{-\Delta G/10}\, p_w \quad \% \quad (4)$$

If the shallow fading range of the distribution is required, follow the method of Step with the following changes: Convert the value of pt obtained to an annual value by using, and use this annual value instead of pt where pt appears. The value of pw calculated is the required annual value p. Following, step If it is required to predict the distribution of enhancement for the average year, follow the method where $A_{0.01}$ is now the fade depth exceeded for 0.01% of the time in the average year. Obtain first pw by inverting and using p □ 0.01%. Then obtain fade depth $A_{0.01}$ exceeded for 0.01% of the time in the average year by inverting as appropriate, and us-ing p in place of pw.

The meteorological data that used in this study has been brought from SMA, Sudan for 10 years from 2004 to 2014 for 5 meteorological stations over the country with 1756 total number of examples. Figure 2 shown general method the used in the study. In addition to , figure 3 shown

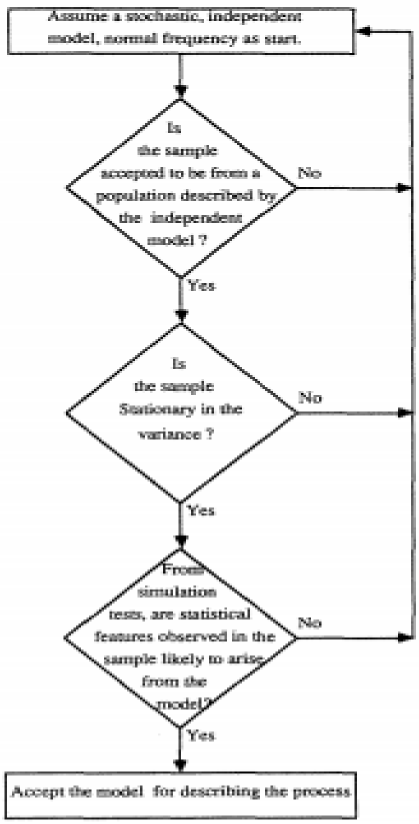
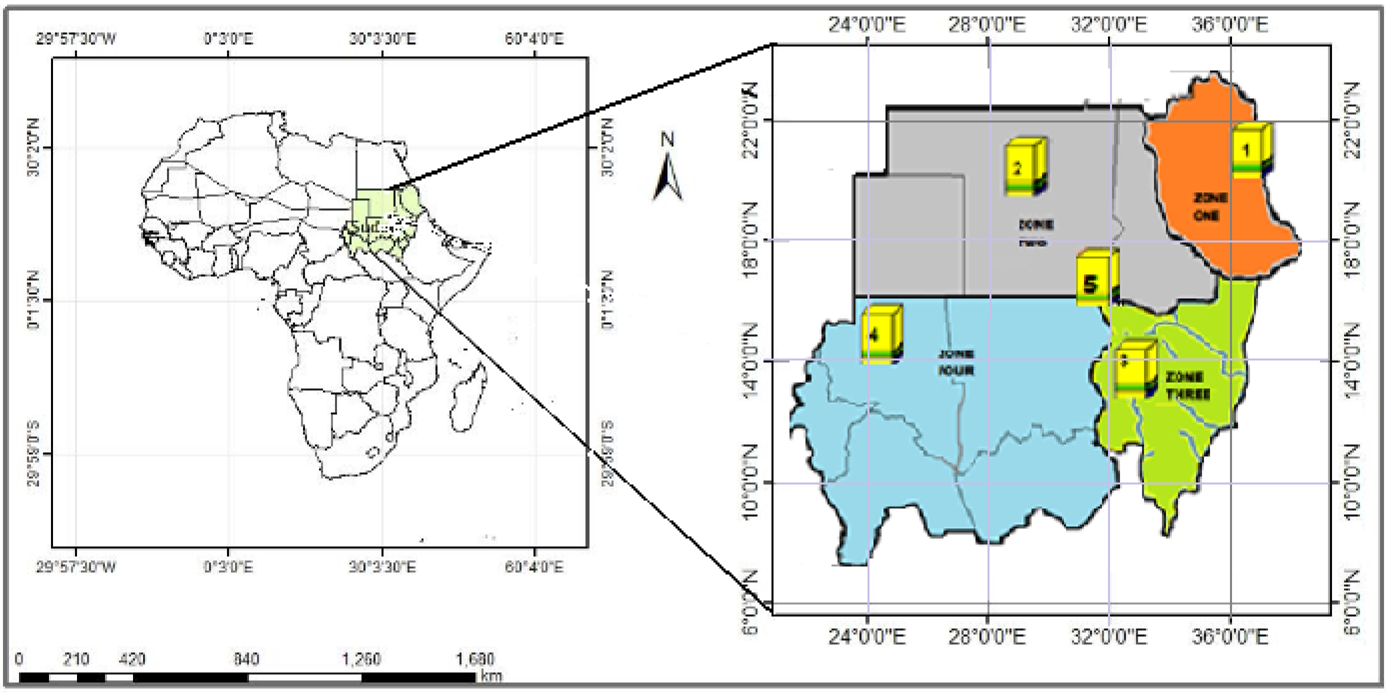


Figure 2: Flow chart of Methodology

Millimeter wave frequencies for ground or satellite links are prone to variations in attenuation over a large range due to rain and dust storm, more than 30 dB some time. Traditionally, it has been considered to be part of meteorological conditions, unavoidable variations. A mathematical model is often an approximate representation of a more complex system. It is important to consider many technical issues before going to design and establish an expensive Wireless system. It is necessary to see the mathematical predictions or calculations of different parameters before going to design such type of systems.

A mathematical model is often an approximate representation of a more complex system. In modeling complex systems, model parameters often abound in number. The value of these parameters may be approximately determined through the fitting of model predictions with calibration data obtained from laboratory experiments, first principle arguments, ab initio calculations, more refined models, etc. Unfortunately, repeating a given experiment multiple times may yield different results that are suitably described by a statistical distribution.



Consequently, model parameters obtained from calibration data sources are themselves often described statistically. This statistical model parameter uncertainty then potentially affects all future predictions that make use of this model. Major task at hand is to propagate this model parameter uncertainty throughout subsequent calculations, thus quantifying the statistical behavior of derived outputs. We refer to this task as uncertainty quantification from statistical model parameter uncertainty.

Figure 3 five 5 meteorological stations of study in Sudan

IV. Results and Discussion

Prediction models for the purpose *of calculating* signal propa*gation and performing fading* predictions are empirical. Empirical models are the result of the application of mathematical regression techniques on measurement data and therefore result in a relationship that describes a variable's dependency under certain given conditions. Empirical prediction models often provide a fair description of the fading process for distances and frequencies that lie within the data range for which measurements have actually been collected. The whole process is iterative and may go through many redesign phases before the required quality and availability are achieved shown in Figure 2 The quality of satellite communication systems can be seriously affected by variable climatic phenomena such as rain and turbulence. For the design of communication *system a* required availability, statistical knowledge of climatic propagation effects is essential. Predicting attenuation due to dust storm is an important but complex problem, and a variety of models have been developed [2]. Different approaches have been adopted by researchers [13] to evaluate microwave and millimeter wave signal attenuation due to sand and dust storm in terms predicting modelling. In addition, the restrained use of MW and MMW bands for commercial operations is due to severe sand and dust storm attenuation. Attenuation experienced in arid and semi- arid region is caused by considerably wind - blown and sand and dust turbulent atmosphere as compared to other parts of the world. Propagation conditions vary from month to month and from year to year, and the probability of occurrence of these conditions may vary by as much as several orders of magnitude. Some phenomena occur only rarely, requiring many years of observation to make any conclusions. For instance, elevated ducting may occur only several times per year in some locations, and in many locations, rain intense enough to affect propagation paths occurs for less than 1 percent of a year. In parts of India at locations where the monsoon occurs some years but not others, several significant rain events may occur one year but not the following year.

Worst-month statistics is of interest to those faced with designing a system to meet performance criteria expressed in terms of a percentage of any calendar month, or of any contiguous 30-day period. The system designer, in this case, needs to find the percentage of time that some threshold value of attenuation or rain rate and dust storm will be exceeded within a given month. For every threshold value, there corresponds a month of the year having the highest percentage of time exceeding the threshold (i.e., the percentage exceedance). This is designated the "worst-month" for that threshold. The percentage exceedance in this month, to be expected once every year or every given number of years, is of most interest. For high rain rates or high visibility in dust storm, the worst-month would probably correspond to the period of highest thunderstorm intensity or frequency whereas the worst-month for lower rain rates might be when most rainfall is of the steady, strati form variety.

V. CONCLUSION

Implications of the results and future research directions are also presented. The main findings from this work can be summarized as follows: various prediction models developed for rain or sand and dust storm are comprehensively described. Secondly, determination of the attenuation statistics is indispensable in planning and designing of wireless communications systems and links. Worst-month statistics of total attenuation were studied by using data collected with 5 radiometers in the Graz area at 3 different frequencies. Reasonable agreement with the Bashir model (6) was found. The results obtained agree quite well with the last year-to-worst-month relation for annual outage probabilities lower than around 0.1% and with for probabilities higher than 1%. However, the data would *seem to* indicate that a constant Q. would be more appropriate in the range1 to around 30%. S ome further study of this *behavior* would be necessary.

ACKNOWLEDGMENT *(Heading 5)*

We wish to thank Razak School of Engineering and Advanced Technology, Universiti Teknologi Malaysia (UTM), Malaysia for providing the facilities, tools, and equipment for the radio signal propagation research.

REFERENCES

[1] S. O. Bashir, *"Statistical Modelling of Propagation Parameters through Sand/Dust Storms at Microwave Frequencies,"* IEEE Intern Conf on Antennas, Propagation and Systems, Johor Bahru Malaysia, 3-5, 2009.

[2] Yagasena and S. I. S. Hassan, "Worst-month rain attenuation statistics for satellite-Earth link design at Ku-band in Malaysia," *TENCON 2000. Proceedings*, Kuala Lumpur, 2000, pp. 122-125 vol.1.

[3] Emiliani, L. D. "Worst-Month Rain Rate and Rain Attenuation Statistics for Satellite System Design in a Mountainous Region in Colombia," in *IEEE Antennas and Wireless Propagation Letters*, vol. 5, no. 1, pp. 475-478, Dec. 2006.

[4] P.A. Owolawi and T.J. Afullo, "Rainfall rate modelling and worst month statistics for millimetric line-of–sight radio links in South Africa," Radio Sci. , vol. 42, 2007.

[5] A. Musa, S. O. Bashir, Md Rafiqul Islam and Othman O. Khalifa "Dust-Storm Induced Cross-Polarization at MMW Bands in Northern Nigeria" International Conference on Computer and Communication Engineering (ICCCE 2012), 3-5 July 2012, Kuala Lumpur, Malaysia

[6] CHEN C, CHEN Y, LIU C, LIN P and CHEN W "Statistics of heavy rainfall occurrences in Taiwan. Weather Forecasting" Vol 22 (5) 981-1002. 2007

[7] M. P. de Villiers and J. van Heerden "*Dust storms and dust at Abu Dhabi international airport* " Royal Meteorological Society, 2007 DOI: 10.1002/wea.42

[8] Teh Sieh Ting and J.S. Mandeep "*Analysis of Worst-month Relationship with Annual Rain Attenuation in Malaysia*" Research Journal of Applied Sciences, Engineering and Technology 7(7): 1453-1455, 2014

[9] Per Thorvaldsen and Ingvar Henne "Long-term propagation measurements on a line-of-sight over-water radio link in Norway"

[10] R. Kumar, M. C. Barth, G. G. Pfister, M. Naja, and G. P. Brasseur "WRF-Chem simulations of a typical pre-monsoon dust storm in northern India: influences on aerosol optical properties and radiation budget" Atmos. Chem. Phys., 14, 2431–2446, 2014.

[11] R. K. Crane, "A local model for the prediction of rain-rate statistics for rain-attenuation models," in *IEEE Transactions on Antennas and Propagation*, vol. 51, no. 9, pp. 2260-2273, 2003.

[12] Ismail, Ahmad Fadzil and Abd Rashid, Nur Emeilin and Din, Jafri and Abdul Rahman, Tharek and Islam, Md. Rafiqul (2012*) Analyses of worst-month rain fade statistics for Ku-band (26GHz) in the tropics.* In: 2012 7th International Conference on Telecommunication Systems, Services, and Applications (TSSA), 30-31 October 2012 , Bali, Indonesia.

[13] M. F. Camara, S. O. Bashir, F. N. M. Isa and A. Musa, "Prediction of Rain-Induced Cross Polarization at Millimeter Wave Bands in Guinea," Computer and Communication Engineering (ICCCE), 2014 International Conference on, Kuala Lumpur, pp. 173-176, 2014.

Dynamic Hybrid Slot Size Bandwidth Allocation Algorithm for Reducing Packet Delay of Real Time Traffic in EPON

Md. Selim Morshed, Monir Hossen, Mohammad Mahbubur Rahman
Department of Electronics and Communication Engineering
Khulna University of Engineering & Technology
Khulna, Bangladesh

Masanori Hanawa
Interdisciplinary Graduate School of Medicine and Engineering
University of Yamanashi
Kofu-shi, Japan

***Abstract*— In the modern data networks, users are communicating via real time traffic very often, i.e., high priority (HP) traffic. The demand is increasing for that interruption free, high speed, high efficiency, best quality of services to perform successful communications. This paper emphasizes on the HP traffic like Video Conferencing, Video Chat, Telemedicine, Teleconferencing, and Interactive games for the real time communication. In contrast, the low priority data is considered as the best effort (BE) traffic because the BE traffic do not require delay sensitive applications. To achieve the best quality data communication for the HP traffic, in this paper, we propose a new dynamic bandwidth allocation (DBA) algorithm called dynamic hybrid slot size bandwidth allocation (DHSSBA) algorithm that is the modified version of the existing hybrid slot-size/rate (HSSR) algorithm. The performance of the proposed scheme has been analyzed by using numerical simulation in terms of the end to end packet delay for both the HP and BE traffic. We also have compared the simulation results of the proposed scheme with the existing HSSR scheme. From the comparison of the simulation results it is clear that the proposed scheme provides less end to end delay for the HP traffic than the HSSR scheme. However, the proposed scheme provides marginally higher delay than the HSSR scheme for the BE traffic at the higher offered loads.**

Keyword: DBA Algorithm, DHSSBA, HSSR, High Priority traffic, Best Effort traffic.

I. INTRODUCTION

With the advancement of communication technology, currently, the optical communication and networking is using for high capacity, high speed and high efficient data transfer. However, the end user optical access network is still in a developing stage. Passive optical network (PON) is a highly capable access network providing numerous advantages to the several service providers without any bandwidth problem [1]. The PON is considered as a promising next-generation access technology to fulfill the high bandwidth demand by the real time data traffic [2]. It is an efficient access network that provides lower operation and maintenance costs and allows larger distance between the central office (CO) and end users [3].

Ethernet PON (EPON) is a most effective standard of the PON technology. The EPON is deployed with bus or tree topology and connects the multiple optical network units (ONUs) with a single optical line terminal (OLT) which eventually connects to the metro and wide area networks (WANs) [4]. A feeder fiber is used to connect the OLT to multiple ONUs and a passive coupler, i.e., splitter/combiner, splits the optical signal into different ONUs in the downstream direction [5]. The EPON does not require any active element between the CO and the end users that only comprises with the passive optical components such as optical fiber, couplers, splices, and splitters [6].

The EPON is a topology based on IEEE 802.3ah standard has 8B/10B encoding with 1.25 Gbps nominal or 1 Gbps data rate effective for both the upstream and downstream directions [6]. The EPON structure with hybrid slot size bandwidth distribution principle is shown in Fig.1. Here, N is the number of ONUs those are connected to the OLT through an optical combiner / splitter. A cycle time is divided into two equal parts, i.e., part 1 is for the high priority (HP) data traffic and part 2 is for the best effort (BE) data traffic. The HP and BE data traffic of each ONU are sent to the OLT by using priority management in the ONU.

One of the main problems of the EPON is to share a common upstream channel by the multiple ONUs. For better

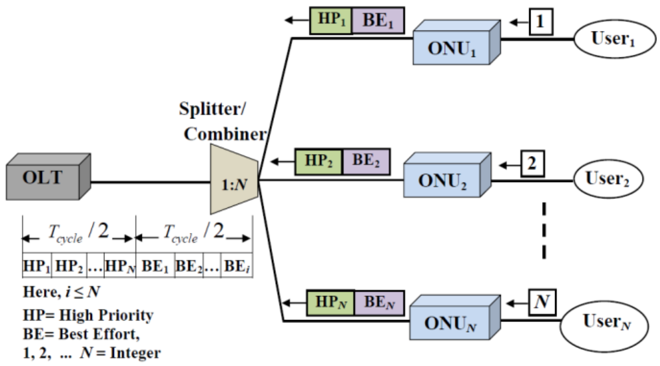


Fig.1. EPON Architecture in proposed method

and efficient upstream data transmission, different dynamic bandwidth allocation (DBA) algorithms have been proposed by different researchers. The overall performance of the EPON depends on the effectiveness of a DBA algorithm.

The interleaved polling with adaptive cycle time (IPACT) is a very basic and popular DBA algorithm which polls every ONUs in a round robin fashion [7]. The IPACT works on a data transmission algorithm called multi-point control protocol (MPCP) that uses two control messages, i.e., GATE from the OLT to the ONUs and REPORT from the ONUs to the OLT, to allocate bandwidth to each ONU for sending data packets in the upstream direction. In the IPACT, the OLT maintains a polling table to store the relevant information, e.g., number of bytes waiting in each ONUs' buffers, round trip time (RTT) between the OLT and ONUs. In each time cycle, the OLT transmits the GATE messages to all the ONUs to inform about the allocated transmission windows in the next time cycle. However, the IPACT is a very basic DBA algorithm that is not effective for the real time or HP data traffic as it does not provide any priority scheduling technique.

Hossen et al. [8] have proposed a new DBA algorithm called adaptive limited DBA (ALDBA) for hybrid PON system where different ONUs are connected to the two different service providers, i.e., fiber to the home (FTTH) and wireless sensor network (WSN). In the ALDBA scheme, two different maximum upstream transmission windows have been considered for the two different service providers as the packet lengths and data rates of the FTTH and WSN are not similar. The ALDBA scheme can effectively mitigate the problem of different data rate and packet lengths of two different service providers but it does not consider the priority based services. That is why, the ALDBA scheme is also not suitable for the real time applications.

In [9], a new DBA algorithm, i.e., hybrid slot-size/rate (HSSR), has been proposed where a single time slot is divided into two parts. The 1st part of a time cycle is used for the HP traffic and in every time cycle a fixed bandwidth is allocated to each ONU's HP data packets. The 2nd part of the time cycle is dynamically allocated for the BE traffic of all ONUs in the network. The HSSR scheme can provide better performance for the HP data traffic than the other conventional DBA algorithms. However, bandwidth of the 1st part of each time cycle will be wasted for the lightly loaded ONUs and that leads lower bandwidth utilization and lower throughput than the conventional DBA algorithms. This bandwidth utilization problem can be solved and overall performance can also be improved by using DBA scheme in the 1st half of the time cycle for the HP traffic.

In this paper, we propose a new DBA algorithm called dynamic hybrid slot size bandwidth allocation (DHSSBA) algorithm that is a modified version of the HSSR scheme. In the proposed scheme, 1st part of a time cycle is dynamically allocated to the HP traffic of all ONUs in the network. Moreover, if the requested window size of the HP traffic of an ONU is larger than the maximum allocated window size for the HP traffic then it will utilize the additional bandwidth from the 2nd part of the time cycle. The main drawback of this proposed scheme is that if the offered load of the HP traffic of an ONU is larger and it occupy additional bandwidth from the 2nd part of the time cycle then the BE traffic will suffer from more delay. The performance of the proposed scheme is evaluated in term of end-to-end packet delay for both cases of the HP and BE traffic. From the comparison of the proposed DHSSBA scheme to the existing HSSR scheme, it is clear that the packet delay of the HP traffic is greatly reduced in the proposed scheme than the existing HSSR scheme. In contrast, the packet delay of the BE traffic in the proposed scheme is slightly higher than the HSSR scheme for the higher offered loads. However, the overall delay, i.e., combined delay of the HP and BE traffic, in the proposed scheme is still lower than the existing HSSR scheme.

The rest of the paper is organized as follows: the section II represents the overview and bandwidth allocation principles of the proposed DBA scheme. In the section III, we explain the simulation results and discussion. Finally, in the section IV, we conclude the paper.

II. PRINCIPLE OF THE PROPOSED DHSSBA ALGORITHM

In this section, we explain about the protocol overview, data transmission principle using the proposed DHSSBA algorithm and data transmission and bandwidth allocation principles.

A. Protocol Overview of the Proposed DHSSBA Scheme

Figs. 2(a), 2(b), and 2(c) show the bandwidth distribution scenarios in different time cycles, i.e., time cycles i, i+1, and i+2, for the conventional, HSSR, and proposed DHSSBA algorithms, respectively.

The conventional DBA schemes of the EPON are slot-size based, where the OLT grants time slot for all the ONUs according to their requested traffic. However, the maximum granted window size of each ONU is upper bounded by the length of a time cycle and number of ONUs in the EPON, i.e., $W^{max} = T_{cycle}/N$, where, T_{cycle} is the length of a time cycle and N is the number of ONUs. In these schemes, the HP traffic and the BE traffic are sent within the same time slot in a time cycle as shown in the Fig. 2(a) here the bandwidth distribution strategy in the time cycle i and i+1 are similar.

In the HSSR scheme [9], each time cycle is divided into two parts where the 1st part is steady, i.e., in every time cycle the granted window size for each ONU is fixed. As shown in the Fig. 2(b), the 1st part of the time cycle is divided into N time slots and each time slot is allocated for the HP traffic of a particular ONU. The time slot length for the HP traffic of an ONU at any time cycle, i.e., time cycle i or i+1 or i+2, is fixed. The 2nd part of each time cycle is dynamic and used for the BE traffic of the ONUs. So, depending on the bandwidth requirement of the BE traffic of all the ONUs in the network the length of the 2nd part of a time cycle is varied.

In our proposed DHSSBA scheme as shown in the Fig. 2(c), primarily each time cycle is divided into two equal parts, i.e., $T_{cycle}/2$. However, for giving more priority to the HP traffic if the bandwidth demand by the HP traffic is larger than

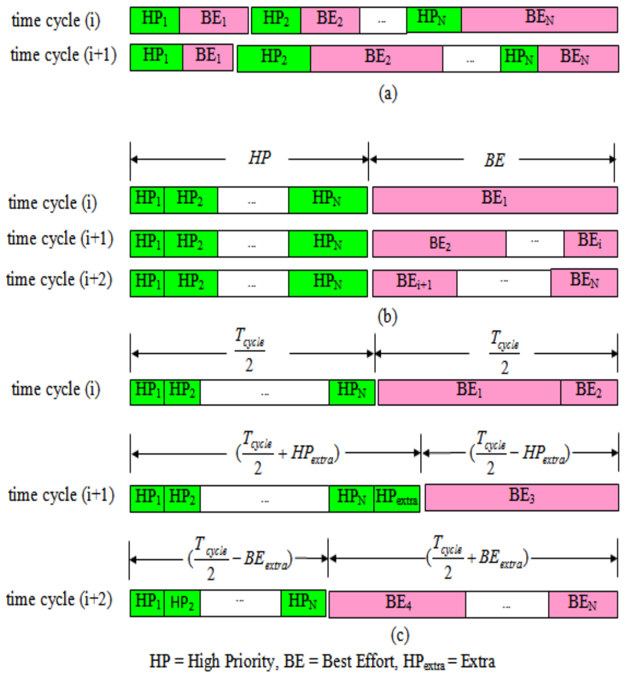


Fig. 2. Time slot allocation diagram in (a) Conventional DBA scheme, (b) HSSR scheme, and (c) Proposed DHSSBA scheme.

the maximum window size then the length of the 1st part of the time cycle can be extended. In the Fig. 2 (c), it is shown that in the time cycle (i+1), the 1st part of the time cycle is extended to the length of ($T_{cycle}/2 + HP_{extra}$). In contrast, the length of the 1st part of a time cycle will be shortened if the bandwidth demand by the HP traffic is lower. In the Fig. 2 (c), it is shown that in the time cycle (i+2), the 1st part of the time cycle is shortened to the length of ($T_{cycle}/2$ - BE_{extra}). That means, in the proposed scheme, the length of a time cycle for the HP traffic is dynamic in nature while in the existing HSSR scheme it is fixed.

The second part of the time cycle is also dynamic and used for the BE traffic for both the proposed and existing schemes. However, one giant slot can be assigned to an ONU's BE traffic if required to reduce the guard intervals. If the network load of BE traffic is low then the 2nd part of the time cycle is shared by the multiple ONU's BE traffic. The main differences in the bandwidth distribution in the 2nd part of the time cycle for the BE traffic between the proposed and existing schemes is that, in the existing HSSR scheme, the maximum length of the 2nd part is fixed, i.e., $T_{cycle}/2$ as shown in the Fig. 2 (b) while, in the proposed scheme, the length of the 2nd part of the time cycle can be increased or decreased depending on the bandwidth demand of the HP traffic in the network, as shown in time cycle (i+1) and (i+2), respectively, of the Fig. 2 (c).

B. Data Transmission Principle in the Proposed Scheme

Fig. 3 shows the data transmission principle for a time cycle in the upstream direction of the proposed DHSSBA scheme. In this figure, the condition is considered that there will be no extra bandwidth demand by the HP traffic of all the ONUs, i.e., the summation of the requested window size by the HP traffic of all the ONUs will be fitted into the 1st part of the time cycle. In contrast, the BE traffic of all the ONUs will be fitted into the 2nd part of the time cycle. From the figure, it is clear that at first the granted window size of the HP traffic of ONUs 1 to N is transmitted to the OLT followed by the bandwidth requests for the next time cycle for both the HP and BE traffic.

C. Bandwidth Allocation Principle in the Proposed Scheme

In the proposed DHSSBA scheme, the maximum transmission windows of an ONU for both the HP and BE traffic are $W_{HP}^{\max} = T_{cycle}^{\max}/(2N)$ and $W_{BE}^{\max} = T_{cycle}^{\max}/(2N)$, respectively. Here, $W_{HP}^{\max}$ is the maximum transmission window for the HP traffic, $W_{BE}^{\max}$ is the maximum transmission window for the BE traffic, $T_{cycle}^{\max}$ is the maximum length of a time cycle and N is the number of ONUs in the network. According to the proposed scheme, the allocated transmission window of an ONU for both the HP and BE traffic are depended on the following three cases:

Case I:

If the overall bandwidth demands, i.e., for both the HP and BE traffic, of all the ONUs is low then the granted transmission windows at a time cycle for both the HP and BE traffic of the ONU i is calculated using (1) and (2):

$$W_{HP}^{grant(i)} = W_{HP}^{R(i)} \quad \text{if } \mathrm{W}_{\mathrm{HP}}^{\mathrm{R(i)}} \le \mathrm{W}_{\mathrm{HP}}^{\max} \tag{1}$$

$$W_{BE}^{grant(i)} = W_{BE}^{R(i)} \quad \text{if } \mathrm{W}_{\mathrm{BE}}^{\mathrm{R(i)}} \le \mathrm{W}_{\mathrm{BE}}^{\max} \tag{2}$$

where, $W_{HP}^{R(i)}$ and $W_{BE}^{R(i)}$ are the requested window sizes of ONU i at a time cycle for the HP and BE traffic, respectively, $W_{HP}^{grant(i)}$ is the granted window size for the HP traffic and $W_{BE}^{grant(i)}$ is the granted window size for the BE traffic of ONU i at a time cycle.

Case II:

If the network is heavily loaded and the requested window size of the HP traffic is larger than the $W_{HP}^{\max}$ then the total excess request of all the ONUs is calculated and excess bandwidth is granted from the 2nd part of the time cycle. In this case, the allocation of the excess bandwidth from the 2nd part of the time cycle does not depends on the conditions of the BE traffic. That means for any case of the BE traffic, either heavily loaded or lightly loaded, the excess bandwidth is allocated to the HP traffic. This principle is strictly maintained to improve the performance of the real time data traffic in the EPON. Following equation is used to calculate the excess bandwidth for the HP traffic of the heavily loaded ONUs.

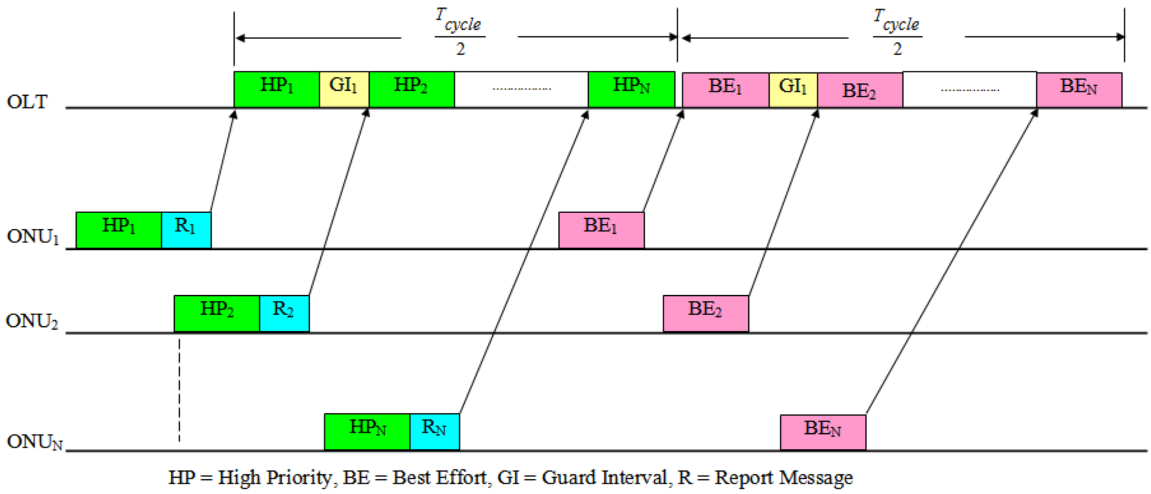


Fig.3. Data transmission diagram in the proposed DHSSBA scheme.

$$W_{HP}^{extra} = \sum_{i=1}^{n} \left(W_{HP}^{R(i)} - W_{HP}^{\max} \right) \tag{3}$$

here, W_{HP}^{extra} is the total extra bandwidth for the HP traffic from the 2nd part of the time cycle and n is the number of heavily loaded ONUs of the HP traffic.

Finally, this W_{HP}^{extra} bandwidth is fairly distributed for the HP traffic among the n heavily loaded ONUs of the network. Following formula is used to allocate the transmission window to the HP traffic of the ONUs in the proposed DHSSBA scheme:

$$W_{HP}^{grant(i)} = \begin{cases} W_{HP}^{R(i)} & \text{if } W_{HP}^{R(i)} \le W_{HP}^{\max} \\ W_{HP}^{R(i)} + \dfrac{W_{HP}^{extra}}{n} & \text{if } W_{HP}^{R(i)} > W_{HP}^{\max} \end{cases} \tag{4}$$

In this case, the granted window for the BE traffic of a heavily loaded ONU can be up to the remaining transmission window, i.e., $\left(T_{cycle}^{\max}/2 - W_{HP}^{extra}\right)$, if required. Otherwise, this remaining transmission window is shared by the BE traffic of multiple ONUs. Following equation is used to grant the transmission windows to the BE traffic of the heavily loaded ONUs in the proposed scheme:

$$W_{BE}^{grant(i)} = \begin{cases} W_{BE}^{R(i)} & \text{if } W_{BE}^{R(i)} < \left(\dfrac{T_{cycle}^{\max}}{2} - W_{HP}^{extra} \right) \\ \dfrac{T_{cycle}^{\max}}{2} - W_{HP}^{extra} & \text{if } W_{BE}^{R(i)} > \left(\dfrac{T_{cycle}^{\max}}{2} - W_{HP}^{extra} \right) \end{cases} \tag{5}$$

Case III:

In this case, the HP traffic of all the ONUs are considered as lightly loaded, i.e., $\sum_{i=1}^{N} W_{HP}^{R(i)} < T_{cycle}/2$, while the BE traffic of all the ONUs are heavily loaded, i.e., $\sum_{i=1}^{N} W_{BE}^{R(i)} > T_{cycle}/2$. As in the (3), the excess bandwidth W_{BE}^{extra} for the BE traffic is calculated using the equation bellow:

$$W_{BE}^{extra} = \sum_{i=1}^{N} (W_{HP}^{\max} - W_{HP}^{R(i)}) \tag{6}$$

This excess bandwidth is added to the 2nd part of the time cycle to provide more bandwidth for the BE traffic of the heavily loaded ONUs. Following equations are used to allocate the transmission window to the heavily loaded ONUs of the BE traffic:

$$W_{BE}^{grant(i)} = \begin{cases} W_{BE}^{R(i)} & \text{if } W_{BE}^{R(i)} \le \left(\dfrac{T_{cycle}}{2} + W_{BE}^{extra} \right) \\ \dfrac{T_{cycle}}{2} + W_{BE}^{extra} & \text{if } W_{BE}^{R(i)} > \left(\dfrac{T_{cycle}}{2} + W_{BE}^{extra} \right) \end{cases} \tag{7}$$

From (7) it is clear that if the BE traffic of an ONU is hugely loaded then a giant window up to the window size of $T_{cycle}/2 + W_{BE}^{extra}$ is granted. However, if the requested window of the BE traffic of ONU i is less than that of the $T_{cycle}/2 + W_{BE}^{extra}$ then the remaining transmission window W_{rem}, i.e., $W_{rem} = (T_{cycle}/2 + W_{BE}^{extra}) - W_{BE}^{R(i)}$, is granted to the BE traffic of the next ONU i+1 as required. The main advantage of this giant window allocation to the BE traffic of an ONU is that it can avoid the bandwidth allocation to the BE traffic of the multiple ONUs at a time cycle that reduces the number of guard time as well as overhead.

III. Simulation Results and Discussion

In this section, the performances of the proposed DHSSBA scheme are evaluated in terms of end to end packet delay for both the HP and BE data traffic. The results of the proposed scheme are compared with the existing HSSR scheme. Computer simulation was used to evaluate the performances of the proposed DHSSBA and existing HSSR schemes. We assumed the PON architecture with one OLT and 16 ONUs in a tree topology based EPON system. We assumed the distance between the OLT and ONUs in the range of 10-20 km and 1 Gbps transmission speed for both the upstream and downstream directions. We also have used the packet length of 1500 bytes [10] for the HP traffic while the double packet length was considered for the BE traffic. The simulation parameters are summarized in Table I .

TABLE I. Simulation Parameters

Symbol	Explanation	Value
N	Number of ONUs	16
T_{cycle}	Length of time cycle	2 ms
D	Distance between the OLT and ONUs	10-20 km
T_G	Guard Time	5 µs
R_U	Transmission speed in both the upstream and downstream directions	1 Gbps
T_E	Length of Ethernet overhead	576 bits
T_R	Length of Report message	304 bits
T_P	Length of a data packet for the HP traffic	1500 bytes
P_{HP}	Maximum number of packets for HP	0 to10

Figs. 4 shows the comparison of average end to end packet delay vs. offered load for the HP data traffic between the proposed DHSSBA and existing HSSR schemes. From the Fig. 4, it is clear that the proposed scheme provides almost constant average delay for the HP traffic for any value of the offered load. In contrast, the average packet delay for the HP traffic for the existing HSSR scheme increases rapidly after the offered load of 0.4. At the lowest offered load of 0.1 the average packet delay of the HP traffic for the existing scheme is 1.5 times larger than the proposed scheme. However, at the highest offered load of 1.0 the average packet delay for the HP traffic of the existing scheme is more than 2.5 times larger than the proposed scheme. From these results it is clear that the proposed scheme is very efficient to maintain less packet delay even for the heavily loaded ONUs.

Fig. 5 shows the maximum delay vs. offered loads between the proposed DHSSBA and existing HSSR schemes. Alike the average delay the proposed scheme also provide constant delay up to the offered load of 0.9 and at this offered load the existing scheme provides 3.5 times more maximum delay than the proposed scheme. At the maximum offered load of 1.0 the maximum delay for the proposed scheme is increased still it is less than that of the existing scheme. So it can be concluded that the proposed scheme provides better delay performance than the existing scheme for both the cases of average and maximum delays.

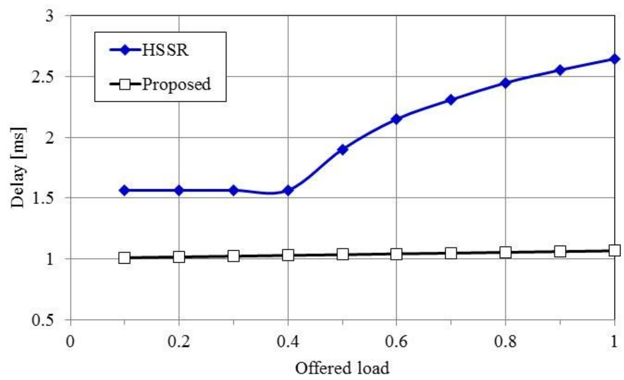


Fig. 4. Average delay in milliseconds for the HP traffic.

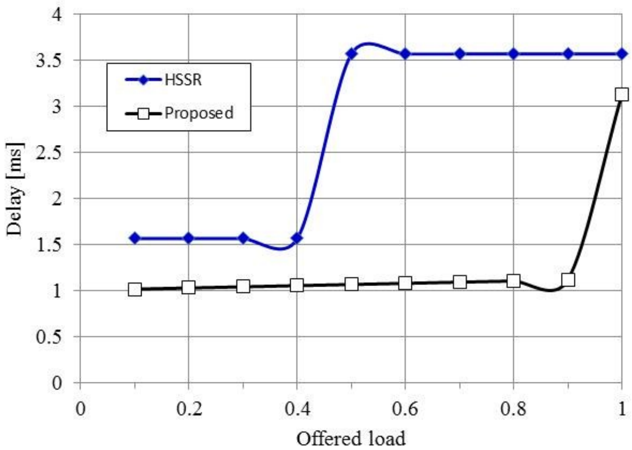


Fig. 5. Maximum packet delay in milliseconds for the HP traffic.

Figs. 6 and 7 show the average end to end delay against offered loads for the BE traffic and overall traffic, i.e., combination of the HP and BE traffic, respectively. From the Fig. 6 it is found that the proposed scheme provides marginally higher delay than the existing scheme at the offered load of more than 0.8. So proposed scheme has some drawbacks in the case of the BE traffic at the higher offered loads. However, from the Fig. 7 it is clear that the overall average delay for the proposed scheme is lower than the existing scheme for any value of the offered load. Therefore, even the proposed scheme provides larger delay than the existing scheme at higher offered loads for the BE traffic the

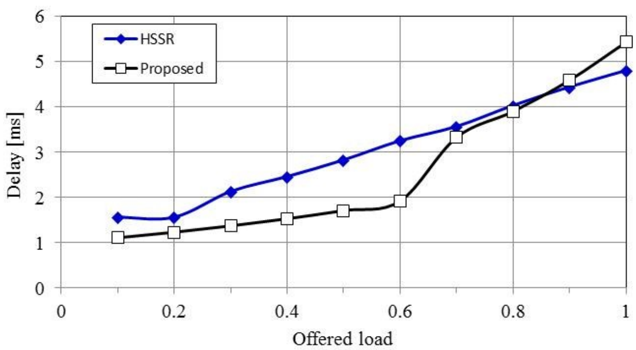


Fig. 6. Average delay in milliseconds for the BE traffic.

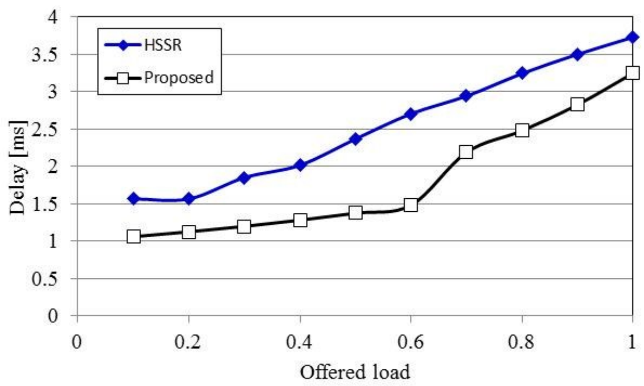


Fig. 7. Average delay in milliseconds for overall traffic (combine traffic of HP and BE).

overall average delay is still lower in the proposed scheme. These results prove that the proposed scheme is more effective than the existing scheme in term of average packet delay for both the HP and combined traffic.

IV. CONCLUSION

In this paper, the main objective of our proposed scheme is to focus on the importance of the transmission of HP data traffic without any losses because the end users never want to sacrifice any real time traffic while communicating with others. Our proposed DHSSBA scheme significantly reduces the average delay than the existing scheme for the HP traffic and also provides constant delay from the very lower value of the offered load to the higher value. At the highest offered load of 1.0 the proposed scheme provides about 60% lower packet delay than the existing scheme for the HP traffic. The main contribution of the proposed scheme is that it can significantly reduce the average packet delay for the HP traffic while the overall packet delay is still lower than that of the existing scheme.

REFERENCES

[1] M. Hossen and M. Hanawa "Dynamic Bandwidth Allocation Algorithm With Proper Guard Time Management Over Multi-OLT PON-Based Hybrid FFH and Wireless Sensor Networks" *IEEE/OSA J. Opt. Commun. Netw.*, Vol. 5, no. 7, pp. 802-812, Jul., 2013.

[2] A. Buttaboni, M. D. Andrade, and M. Tornatore "A Multi-Threaded Dynamic Bandwidth and Wavelength Allocation Scheme With Void Filling for Long Reach WDM/TDM PONs" *IEEE J. lightwave technol*., Vol. 31, no. 8, pp. 1149-1156, April, 2013.

[3] L. Zhou, N. H. Wong, Y. Yeo, X. Cheng, X. Shao, Z. Xu "Traffic scheduling in hybrid WDM-TDM PON with wavelength-reuse ONUs" *J. Photon Netw. Commun.* Vol. 24, pp. 151-159, 2012.

[4] F. Aurzada, M. Scheutzow, M. Herzog, M. Maier, and M. Reisslein " Delay analysis of Ethernet passive optical networks with gated service" " *J. Opt. Netw.,* vol. 7, no. 1, pp. 25–41, Dec. 2007.

[5] B. Kantarci, Member, IEEE, and H. T. Mouftah, Fellow, IEEE "Bandwidth Distribution Solutions for Performance Enhancement in Long-Reach Passive Optical Networks" *IEEE Commun. Surveys & Tutorials,* Vol. 14, No. 3, Third Quarter, 2012.

[6] N. Radzi, N. Din, M. Al-Mansoori, S. Sadon, "Recent Dynamic Bandwidth Allocation Algorithm Methods in Ethernet Passive Optical Network," *International J. of New Computer architectures and their Applications (IJNCAA)* Vol. 4, no. 4, pp. 167-176, 2014.

[7] G. Kramer, and B. Mukherjee, "IPACT: A dynamic protocol for Ethernet PON (EPON)," *IEEE Commun. Magazine*, vol. 40, no. 2, pp. 74-80, Feb. 2002.

[8] M. Hossen, and M. Hanawa, "Adaptive Limited Dynamic Bandwidth Allocation Scheme to Improve Bandwidth Sharing Efficiency in Hybrid PON Combining FTTH and Wireless Sensor Networks" *IEICE Transaction on Commun.*, Vol. E96-B, no.1, pp.127-134, Jan. 2013.

[9] F. An, Y. Hsueh, K. S. Kim, I. M. White, and L. G. Kazovsky, "A new Dynamic Bandwidth Allocation Protocol with Quality of Service in Ethernet based Passive Optical Networks." *Proc. IASTED Int'l. Conf. Wireless and Optical Communications (WOC 2003)*, vol. 3, pp. 165–69, July 2003.

[10] B. Lannoo, L. Verslegers, D. Colle, M. Pikavet, P. Demeester, and M. Gagnaire, "Thorough analysis of the IPACT dynamic bandwidth allocation algorithm for EPONs," Proc. 4th *International Conference on Broadband Communications, Networks and Systems*, pp.486-494, 2007.

A Systematic analysis on the Telemedicine Services in Bangladesh

Uzzal Kumar Prodhan
Computer Science & Engineering Department
Jatiya Kabi Kazi Nazrul Islam University
Trishal, Maymensingh, Bangladesh
uzzal_bagerhat@yahoo.com

Muhammad Zahidur Rahman, Israt Jahan
Computer Science & Engineering Department
Jahangirnagar University
Savar, Dhaka, Bangladesh

***Abstract* – Bangladesh is small country with huge number of populations. The health system of Bangladesh is divided into mainly two sections: Urban and rural based. Most of the expert doctors are urban based with all the modern facilities. But for the poor people of Bangladesh, limited healthcare facilities are available. In this scenario, Telemedicine is the proper method to serve this huge number of peoples. This paper analyzes the Telemedicine models of Bangladesh, finds the potential difficulties faced by the models and proposes some standard recommendations for the Telemedicine in Bangladesh. Analysis of different Telemedicine models include the working procedure of the models, essential features of the models, block diagram based analysis and activity diagram of the models. Finally, authors believe that the findings of this study can be used for the development of Telemedicine in Bangladesh.**

Keywords—telemedicine; crp; dghs; dab; sndp

I. INTRODUCTION

The healthcare facilities are one of the essential basic needs for the people of Bangladesh. The Ministry of Health and Family Welfare are the government body for the people related to their healthcare facilities, enhancement of initiatives, policy development and related activities. The ratio of patients and doctors are very high in Bangladesh. As a result, peoples are not getting proper healthcare services from doctors. In rural areas there are huge shortages of doctors. In Bangladesh, first Telemedicine services were started by Center for Rehabilitation of Paralyzed (CRP), Bangladesh in 1999 to deliver expert doctors opinion to local patients. This initiative is the pioneer telemedicine projects in Bangladesh. All other projects started upon the concepts on CRP projects. Time and cost can be reduced through Telemedicine consultation for rural patients [1], [2], [3]. Medical care facilities can be improved by using telemedicine services in rural [4], [5]. Tele-consultant can affect identification and cure in telemedicine [6].

II. MATERIALS AND METHODS

In order to carry out this study, the different Telemedicine initiatives of Bangladesh are searched from 1990 to 2015. After searching through Journal papers, conferences, books, magazines, related sites and databases the findings are given in the Results sections. The detail finding on the models and recommendations are given in the Discussions and Recommendations section. The summary of the study of Telemedicine projects of Bangladesh are given in the Conclusion section.

III. RESULTS

The first Telemedicine project in Bangladesh was found in 1999 [7]. This project was initiated by Center for Rehabilitation of Paralyzed (CRP), Savar, Dhaka, Bangladesh.

A. Model-1:

Center for Rehabilitation of Paralyzed is basically for spinal injured patients in Bangladesh. Telemedicine was made possible with the support of the Swinfen Charitable Trust in the UK. At present, 9 additional CRP sub-centers across Bangladesh are running by CRP and high quality services now given to persons with disabilities. CRP offers Telemedicine services for the both in-house and outside patients from 1999. The Telemedicine service is free for the in-house patients. The patients can get the expert doctors advice from UK with free of cost. The outdoor patients will have to pay a Telemedicine fees of Taka 1000 for getting this service. CRP uses telemedicine for patients who require a second opinion from consultants in the UK. CRP have access to consultants in the UK with a variety of specialties including from the Royal Hospital, Haslar, who very kindly agreed to provide consultation free of cost. Consultants within Bangladesh are invited to provide and receive Telemedical consultation for their patients through the telemedicine link at CRP.

Spinal cord injuries in-patients get 24 hours a day treatment from CRP. Treatment for the out-patients are open from 8.00 am to 5.00 pm expect Thursday. From the financial year report of 2013-14, it was found that 58,687 patients received out-patient treatment services from CRP [8].The flowchart for the CRP hospital telemedicine activities are shown in the figure-1:

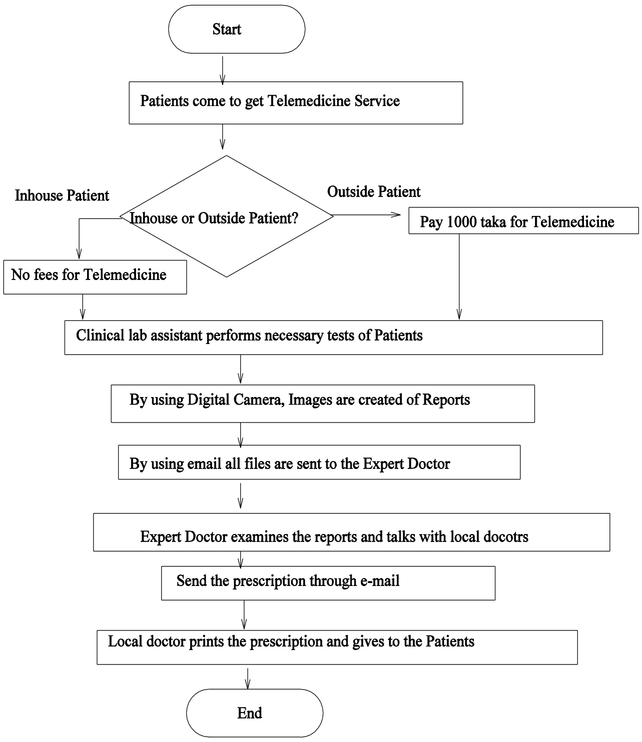


Fig. 1. Flowchart of CRP Telemedicine Service

From the flowchart of the Telemedicine activities of CRP, we can see that the technology used in this project is store and forward based. Email based consultation was the main service category for the patients. It was also helpful for the local doctors of CRP to the exchange of their knowledge with the foreign doctors for any special cases.

B. Model-2:

Telemedicine Reference Center Limited (TRCL) started telemedicine project in the US Trade Show 2001 in Dhaka by using Icare software and normal internet connection. It started test-run of the system between the physicians of Bangladeshi and United States [7], [9].

C. Model-3:

Sustainable Development Network Program (SDNP) Bangladesh has four nodes which were started in January 2003. These nodes are connected through VSAT to satellite. TRCL can act both as service provider and solution provider. Consultancy and diagnostic support are given to the physician at the remote end through medical experts at the SDNP head office [10].

D. Model-4:

Store and forward based telemedicine service was started by Bangladesh University of Engineering & Technology (BUET) and Comfort Nursing Home in 2003. This pilot project continued for a certain period but due to some difficulties the project is not running [7], [9].

E. Model-5:

A pilot telemedicine project was launched by Bangladesh DNS diagnoses Centre, Gulshan-1 and Comfort Diagnoses & Nursing Home in 2004. This pilot project was stopped for the lack of proper publicity, financial and patient -doctors disinterest [7], [9].

F. Model-6:

In order to access to quality healthcare and specialist doctors of Bangladesh from rural areas Diabetic Association of Bangladesh (DAB), Dhaka and Faridpur General Hospital started the pilot telemedicine project with the collaboration of Grameen Telecom in 2005. It was a real time consultation project. In some cases video conferencing system did not work properly, the quality of the images was not satisfactory, lack of dedicated bandwidth and finally the interest of the patients were not good [11], [12].

G. Model-7:

Mobile Maternal, Newborn, and Child Health (MNCH) and Telemedicine Services were started by BRAC, a leading Non Government Organization in Bangladesh for slum people. To improve the health status of the slum population, particularly women and children, BRAC initiated Manoshi project in 2007 at urban slums of nine city corporations around Bangladesh [13]. At present, this project is serving eleven city corporations. The Activity diagram of BRAC health project named Manoshi (MNCH Urban) is shown in the below Figure-2:

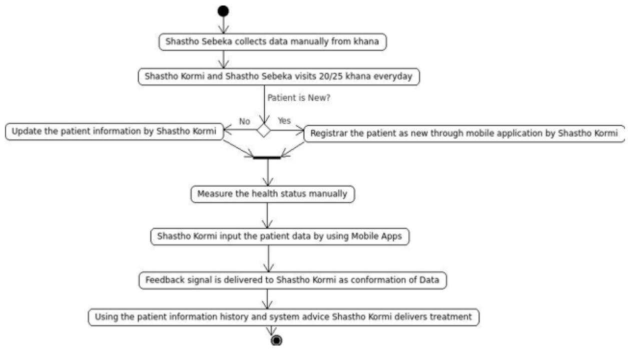


Fig. 2. Activity diagram of BRAC health project "Manoshi"

From the activity diagram, we can find the process of working the project Manoshi (MNCH urban) of BRAC. Shastho Sebeka and Shastho Kormi are the essential component of this project. Through them BRAC collects the slum people health data and store in the central server operated and maintained by BRAC for further health service delivery.

Mobile phone based telemedicine services are also offered to the slum peoples of Bangladesh through BRAC. In every delivery center, there are paramedics who give the health services to the slum peoples. Users can also get Telemedicine services through call center where there are panels of doctors whose are ready to address the patients problem. Doctors give the suggestions or treatment to the caller after hearing the disease details. For fruitful Telemedicine consultation, a conference is also arranged with the Shastho Kormis or Program Organizer and Doctor for the treatment of the patients. Hospitalization is also made to the nearest one through this conference, if the patient's condition becomes critical. The Telemedicine services offered by BRAC are shown in the activity diagram of Figure-3:

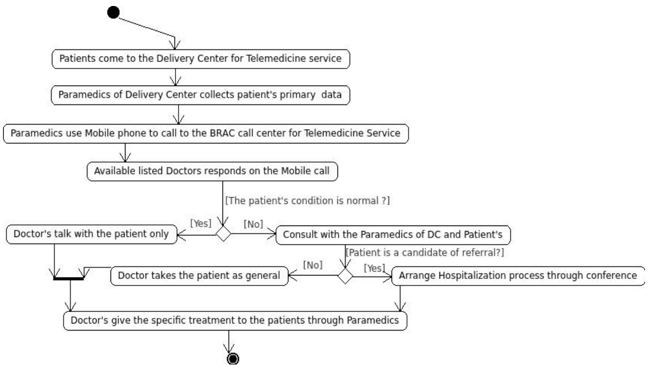


Fig. 3. Telemedicine services offered by BRAC

H. Model-8:

Medinova hospital is operating Telemedicine service from 2006 to till now [7]. This hospital has MoU with the Apollo Hospitals India for offering Telemedicine services to the people of Bangladesh. Doctors from different categories of Medinova Hospitals are engaged with the Apollo hospitals for Telemedicine service. The second or third visit with the expert doctors are made through the Telemedicine services offered by Medinova hospitals for the first time visit patients in India. The flowchart of Medinova hospital Telemedicine service is depicted in the following figure-4:

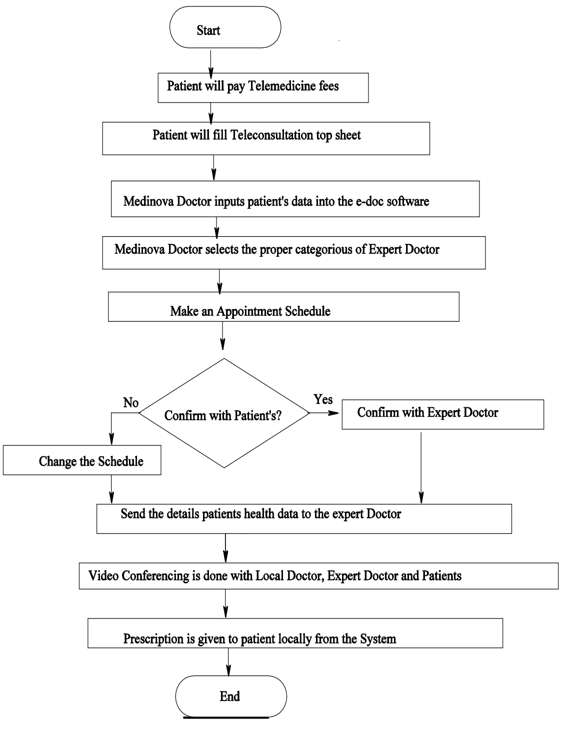


Fig. 4. Flowchart of Telemedicine Service in Medinova Hospital

The flowchart of Medinova hospital shows the overall working activities of Medinova hospital telemedicine process. Patients will have to pay USD 1000 as the fees for the service. Through this service, a patient now is getting the services from the Indian Apollo hospitals doctors without going to India.

I. Model-9:

Grameenphone limited launched a 24 hours health line service to all of its subscribers called "HealthLine Dial 789". Subscribers can make a voice and video call to the licensed physician at the time of their emergency and non-emergency time. An SMS based report delivery service is also attached with this service [14].

J. Model-10:

Ministry of Health, Bangladesh implemented mobile phone based health services for the general people of Bangladesh in 2009. One mobile phone was given to all Upazilla and District level hospital doctors for the use of mhealth service. It was open 24 hours and 7 days. Citizens can make a phone call to the doctors of Upaziila and District hospitals and get the medical advice from them [15].

K. Model-11:

Telemedicine Service in Union Information & Service Centers was introduced by Access to Information project, Bangladesh in 2010. Expert doctors sitting in the MIS office delivering the health advices to UISCs every day. Through videoconferencing technology, rural people are getting medical care without going to the urban areas [16].

L. Model-12:

Aponjon mHealth service was started in Bangladesh by USAID in 2011 with a view to contribute to reduction in maternal and neonatal mortality. Now Aponjon is a health information service using mobile phones to improve health outcomes in Bangladesh. This service was launched in national wide in 2013 in Bangladesh. In order to give service to its receiver, customer service center is introduced for the subscribers and to entry subscribers, answer to the queries of the people and receive complains [17].

M. Model-13:

Grameen Phone in corporation with the Ministry of ICT, Bangladesh has started the Telemedicine services at Jessore District in 2013. Peoples of Jessore are getting the health services with the payment of only BDT 200 or 300 from the expert doctors. This process reduces the travel cost and time for patients [18].

N. Model-14:

Government of Bangladesh has implemented Telemedicine services with the help of Ministry of Health and Family Welfare. Initially the service was limited to 8 Government specialized hospitals only [19]. At present 78 hospitals are operating this project. But Government of Bangladesh will expand this project phase wise at the Upazilla level. All Upazilla health complexes will be under the coverage of telemedicine services. In this process, health services will be available to the rural people of Bangladesh. At present Government Medical Colleges, District Hospitals and Specialized hospitals are the Telemedicine service providers and selected Upazilla health complexes are the Telemedicine service receivers. The hardware organization of this project is shown in the figure 5:

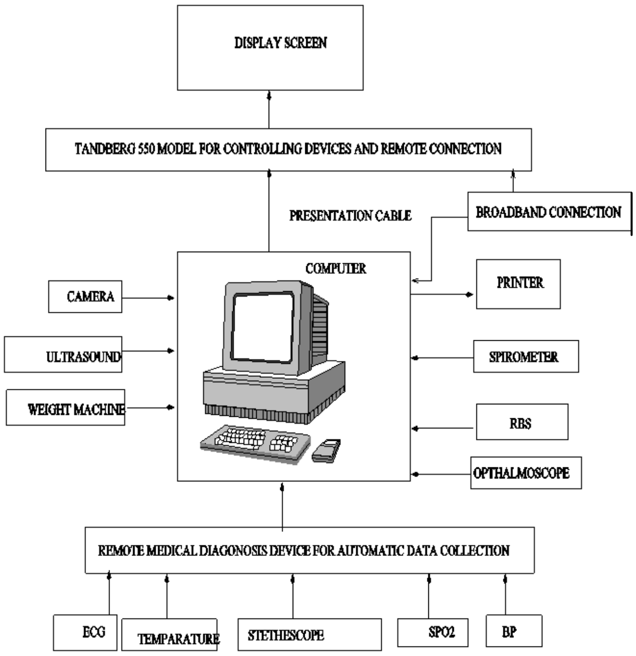


Fig. 5. Hardware Organization of a Government Telemedicine Center

When any outdoor patient is come to the hospital for telemedicine services, he/she will have to buy a ticket from ticket counter with the payment of only 3 taka. Doctors referred the telemedicine patients to the Telemedicine center for the expert opinion. The service engineer checks the doctor's referral documents and selects the patient's disease category. He has a Telemedicine service schedule prepared previously from the corresponding service provider. After examining the schedule, a date and time is fixed for the consultation. At the time of consultation, local hospital doctor, patient, service engineer and expert doctors are present in the conference. Local doctors collect the patient's history manually and record the other health information. There is a Remote Medical Diagnostic device which is used at the time of consultation to send the patient information directly to the expert doctors. At present five vital human body signals like ECG signal, Temperature, Stethoscope, BP and SPO2signal can be sent and monitored both in service provider and service receiver's monitor for live presentation. After a live discussion with the patient and others, a prescription is prepared in the remote center which is given to the patient. The prescription is saved with the patient ID and token number. This ID and token number is used in future for the findings of the history of the patients.

IV. DISCUSSIONS AND RECOMMENDATIONS

This paper presents and analyzes the Telemedicine initiative of Bangladesh. From the analysis, we came to know that most of the initiatives were in project basis. But the present Government of Bangladesh took the Telemedicine project as a priority basis and enhancing the project phase wise. The details finding are given in the following Table I and Table II:

From the study conducted by Lal B Rawal and others, we came to know that no extra benefits are given for health workers who work in rural areas. As a result, no one wants to stay in rural areas for a long time and satisfactorily engage in the treatment of huge number of rural patients [20]. Telemedicine can play a vital role in these types of cases.

TABLE I: SUMMARY OF THE FIRST SEVEN TELEMEDICINE PROJECTS OF BANGLADESH

Model Name	Findings
CRP	1. Uses store and forward based Telemedicine 2. Digital camera is used for capturing images 3. No real time technology is used 4. Indoor patients are free for Telemedicine and outdoor patients need 1000 taka for consultation. 5. No storage for patient data.
TRCL	1. Telemedicine services between Bangladeshi physicians and United States. 2. Used Icare software 3. Internet is used for the service.
SNDP	1. VSAT is used for this project 2. Connected different nodes of Bangladesh. 3. Different consultancy and diagnostic support are given
BUET and Comfort Nursing Home	1. Uses store and forward based Telemedicine 2. This project was running for short period of time.
Bangladesh DNS diagnoses Centre, and Comfort Diagnoses & Nursing Home	1. This project was discontinuous for financial reasons 2. Lack of promotional activities found 3. Patient disinterest also observed
DAB and Faridpur General Hospital	1. Hardware failure was found in this project 2. Internet connection was disrupted 3. Power failure was frequent 4. Performance of the camera used was not satisfactory 5. Radio link was used for Internet connectivity 6. 600 taka was the fees for consultation
BRAC MNCH	1. Slum people are the beneficiary of this project 2. Patient data is stored in the BRAC data center for further use and treatment plan 3. Limited cost for the service 4. Android phone is used for the information entry and update

TABLE II: SUMMARY OF THE LAST SEVEN TELEMEDICINE PROJECTS OF BANGLADESH

Model Name	Findings
Medinova Hospital	1. Use customized software for the storage of patient data and management 2. The cost of the service is $1000 3. Real time video conferencing is used in this project
HealthLine Dial 789	1. Grameenphone used this service as Mobile based 2. SMS based lab report delivery service is also used
mHealth by Ministry of Health	1. Mobile phone is used as the consultancy purpose of District and Upazilla level people at any time round the clock.
Access to Information project	1. Laptop is used for Telemedicine 2. Wireless modem is used for Internet connection 3. Video conferencing is used in this project
Aponjon mHealth	1. This service is mhealth based 2. This service is now Nationwide in Bangladesh
Ministry of ICT and Grameenphone project at Jessore	1. TIMES and DICOT is used for consultancy and record keeping purpose 2. Taka 200 to 300 is required for the registration of patients to get services from this project
Ministry of Health and Family Welfare	1. Largest telemedicine service provider in Bangladesh 2. 78 centers are now running through this project 3. Real time technology is now used 4. No central storage for patients data 5. Hardware failure is seen in some centers 6. Connectivity and power failure is frequent in remote areas

From the findings of the Telemedicine models of Bangladesh, we can conclude that different initiatives were taken for the development of Telemedicine in Bangladesh. The initiatives taken from 1990's to early 2000 were all store and forward basis. This process was time consuming and patients will have to wait for the treatment. The recent initiatives from both Government and Private hospitals are real time based through video conferencing.

The most challenging issues for the Telemedicine projects are connectivity issue. Stable internet connectivity is must for smooth telemedicine. Minimum 2 Mbps dedicated internet connectivity is must for Telemedicine. But these requirements are not ensured in most of the projects. Sometimes internet connection was disrupted at the time of consultation. Considering this, Bangladesh Government is expanding the country wide internet connectivity to all union level.

The study conducted by Emmanuel Kwame Darkwa and others in 2015 shows that the living status in rural areas is very poor. High regulations of Doctors and Nurses are found in rural areas for the lack of facilities. Most of the Doctors are urban based. As a result, healthcare system of rural areas is greatly hampered in Bangladesh [21]. Considering the recent study, this paper recommends to give allowance to the rural health professionals whose are involved in the Telemedicine system. We believe that this system makes them more attentive and effective for the delivery of health care.

From the paper written by Sherwin I. DeSouza and others in 2014, we can find that 99 % respondents want to receive health-related information on their mobile phones. This survey indicated that the mobile phone is used for communicating health information in India [22].

In most of the Telemedicine projects implemented in Bangladesh are lack of data storage facilities and there are no standard formats of patient data. Different models use different format of data. As a result, interoperability of data is not possible within hospitals. So our study suggests a uniform format patient data for the global use. A policy should be developed under the guideline of the experts from both Government and Non Government organizations.

The hardware used in Telemedicine projects is different from one model to another. Most of the models use analog hardware to measure vital information of patients. Government projects use Remote Medical Diagnosis Instrument to measure maximum five vital signs. So our study recommend to develop a different tools to measure the vital human signals which can be easily transferred to remote doctors and database for the quick delivery of patient treatment.

From our study, we can find that the operating cost of Telemedicine is also different. Most of our rural people are very poor. They are now living under the poverty line. So our study recommends fixing this cost minimum so that poor people can afford the cost and use the system.

Most of the Telemedicine projects in Bangladesh work well for a certain period of time. Our study found that the lack of publicity and users acceptance is the main reason for the failure of these projects. So this paper suggests for taking proper steps and necessary publicity policy for the wide use of the Telemedicine services in our country.

At present Government Telemedicine project that is running in the government hospitals are the largest projects in Bangladesh. According to the plan of Government, this project will be expanded to all the Upazilla level. Our study suggests of storing the patient data in a central database which can be accessed from any platform by following a standard protocol. The vital human body signal collections tools should be developed in such a way to collect more vital signs at a time in a cost effective way.

V. CONCLUSION

This paper studies the past and ongoing telemedicine services of Bangladesh. From the analysis of the models, we have shown the important characteristics of the projects. The difficulties of the models are critically found and the recommendations are given based on them. The government of Bangladesh has taken massive plan to introduce Telemedicine in both private and public sector for the rural people of Bangladesh. Authors expect that the results and the proposed recommendations of the study will help the researchers and policy makers for the further development of standard telemedicine model in Bangladesh.

ACKNOWLEDEGEMENT

The authors would like to thank to Dr. Asraf Ali, Associate Professor, Software Engineering Department, Daffodil International University, Dhaka, Bangladesh for his valuable suggestions in order to format this paper. Authors also thank to the Ministry of Information and Communication Technology, Bangladesh for their financial support for the PhD fellowship.

REFERENCES

[1] H.M. Kim, J.C.Lowery, J.B. Hamill and E.G.Wilkins, "Accuracy of a web based system for monitoring chronic wound", Telemedicine Journal and e-Health, vol. 9, no. 2, pp. 129-140, 2003.

[2] M. K. Dobke,A. Renkielska, J. De Neve,J. Chao and D. Bhavsar, "Telemedicine for problematic wound management: enhancing communication between long term care, skilled nursing, and home caregivers and a surgical wound specialist", Wounds, vol. 18, no. 9, pp. 256-261, 2006.

[3] M.K Dobke, D.Bhavsar, A.Gosman, J.De Neve, and B.De Neve, Streamlining the management of patients with problematic wounds: must a multidisciplinary team formulate all patients management plans?", Wounds, vol. 19, no. 12, pp. 340-344, 2007.

[4] A.Tsilimigaki, S.Manaka and T.Tesekoura, "Eighteen months experience with remote diagnosis, management and education in congenital heart disease", Journal of Telemedicine and Telecare, vol. 7, no. 4, pp. 239-243, 2001.

[5] M.J. Kiuru, T.A Paakkala,T.T. Kallio, J.Aalto and M.Rajamaki, "Effect of teleradiology on the diagnosis, treatment and prognosis of patients in a primary care centre", Journal of Telemedicine and Telecare, vol. 8, no. 1, pp. 25-31, 2001.

[6] Ann B.Bynum and Cathy A.Irwin, "Evaluation of the effects of consultation characteristics on telemedicine diagonosis and treatment", International Journal of Telemedicine and Applications, vol. 2011, no. 3 , 2011.

[7] M. Sanaullah Chowdhury, Md. Humayun Kabir, Kazi Ashrafuzzaman and Kyung-Sup Kwak "A telecommunication network architecture for telemedicine in Bangladesh and its applicability," International Journal of Digital Contents Technology and its Applications, vol. 3, no. 3, September 2009.

[8] CRP (2013). "Annual Report 2013-14" Available [online] FTP: http://www.crp-bangladesh.org/index.php?option=com_joomdoc&task=cat_view&gid=41&Itemid=65

[9] Ahsannun Nessa, Moshaddique Al Ameen, Sana Ullah and Kyung Kwak "Applicability of telemedicine in Bangladesh: current status and future prospects," International Arab Journal of Information Technology, vol. 7, no. 2, Page 138-145, April 2010.

[10] Michael j. Crawley, "The R Book" 2nd ed., John Wiley & Sons, Ltd., 2013.

[11] Vassallo D., Hoque F., Roberts M., Patterson V., Swinfen P., and Swinfen R., "An evaluation of the first year's experience with a low cost telemedicine link in Bangladesh," Computer Journal of Telemedicine and Telecare, vol. 7, no. 3, pp. 125,138, 2001.

[12] Istiak M. and Shahriar K., " Telemedicine Services at the Diabetic Association of Bangladesh with Technical Collaboration of Grameen Telecom", The proceedings of Prospect and Problem of Mobile and Land Phones, no. 1, pp. 266-268, 2002.

[13] Tapash Roy, Lucy Marcil, Rashed H. Chowdhury, Kaosar Afsana and Henry Perry, "The BRAC Manoshi Approach",2007 Available[online]. FTP: http://brac.net/health-nutrition-population/item/867-manoshi

[14] Md. Mahfuz Ashrafi, Noushin Laila Ansari, Bushra Tahseen Malik and Barnaly Rashid, "Evaluating the impact of mobile phone based health help line" service in rural Bangladesh", 2011. Available [online]. FTP: http://www.globalhealthhub.org/2011/07/11/evaluating-the-impact-of-mobile-phone-based-%E2%80%98health-help-line%E2%80%99-service/N.Aoki, K.Dunn, K.A.Johnson-Throop and J.P.Turley, "Outcomes and Methods in Telemedicine Evaluation", vol. 9, no. 4, pp. 393-401, 2003.

[15] Tanvir Ahmed, Gerald Bloom, Mohammad Iqbal, Henry Lucas, Sabrina Rasheed, Linda Waldman, Azfar Sadun Khan, Rubana Islam and Abbas Bhuiya, "E-health and M-Health in Bangladesh: Opportunities and Challenges", March 2014.

[16] DGHS (2013). "Telemedicine Service in Union Information and Service Centers", Available [online]. FTP: http://www.dghs.gov.bd/index.php/en/e-health/our-ehealth-eservices/84-english-root/ehealth-eservice/99-telemedicine-service-in-union-information-service-centers

[17] Radha Rajan and Alain Labrique, "MAMA Aponjon: Translating research into mHealth practice", 2014, Available [online]. FTP: http://www.mobilemamaalliance.org/node/745

[18] Abdullah Mamun (2013). "GP launches Telemedicine services in Jessore", 2013, Available [online]. FTP: http://archive.thedailystar.net/beta2/news/gp-launches-telemedicine-service-in-jessore/

[19] DGHS, "Telemedicine Service", Available [online]. FTP: http://www.dghs.gov.bd/index.php/en/component/content/article/84-english-root/ehealth-eservice/490-telemedicine-service

[20] Lal B Rawal, Taufique Joarder, Sheikh Md. Shariful Islam, Aftab Uddin and Syed Masud Ahmed "Developing effective policy strategies to retain health workers in rural Bangladesh: a policy analysis", Human Resources for Health, 13:36, 2015.

[21] Emmanuel Kwame Darkwa, M. Sophia Newman, Mahmud Kawkab and Mahbub Elahi Chowdhury "A qualitative study of factors influencing retention of doctors and nurses at rural healthcare facilities in Bangladesh" BMC Health Services Research, 15:344, 2015.

[22] Sherwin I. DeSouza , M. R. Rashmi, Agalya P. Vasanthi, Suchitha Maria Joseph and Rashmi Rodrigues "Mobile Phones: The Next Step towards Healthcare Delivery in Rural India?",PLOS ONE, 2014.

Selection of Optimal Number of Relays for Distributed Wireless Networks Based on Game Theory

Abhijit Chowdhury
Dept. of Computer Science & Engineering
Chittagong University of Engineering & Technology
Chittagong, Bangladesh
Chowdhury.abhi89@gmail.com

Asaduzzaman
Dept. of Computer Science & Engineering
Chittagong University of Engineering & Technology
Chittagong, Bangladesh
asad@cuet.ac.bd

***Abstract*—Relay nodes can help the source node to transmit information towards the destination. In a distributed communication system, finding out the optimal number of relay is a critical task. Applying Game Theory in distributed wireless network has advantage in efficient relay selection. In this paper, we propose a relay selection method for a distributed multi-node network, based on the power allocation by using game theory without the knowledge of Channel State Information (CSI). The distributed resource allocation only needs local knowledge (bandwidth and power) about channel condition. We are using Stackelberg game for cooperative environment and Subgame Perfect Nash Equlibrium (SPNE) for non-cooperative environment to obtain near-optimal solution by less computation cost and overhead. The decision of near optimal solution is taken based on the value of utility function. Simulation result represents that, relay nodes which are located between the source and destination node and relatively closer to source node offerthe better transmission rate. Thus we obtain the location range of the best relays and the number of relay nodes for which the utility value is maximized.**

***IndexTerms*—Relay selection, Distributed wireless network, Stackelberg game theory, Subgame perfect Nash Equilibrium.**

1. INTRODUCTION

Cooperative devices and mechanisms are increasingly important to enhance the performance of wireless communications and networks. In cooperative communication multiple nodes work together to form a virtual antenna and the relay nodes play a significant role for achieving better performance of the network. But all the relay nodes cannot make same contribution in a wireless network. Find out the optimal number of relay nodes and the location of the suitable relays for a distributed wireless communication is a challenging task.

In a multiple relay system, if any relay is dead or in deep fade the receiver can still get data from other relays [1]. The work in [2] describes, each user transmits its own data towards the base station and also serves as a relay for other users. According to the authors of [3] a single "best" intermediate node is selected by means of partial relay selection (PRS) technique to transmit the information to the destination while the nodes are using AF. The use of multiple relays in a network consists of single source and multiple destinations, brings the issue of how best to assign the relays to the destinations. Optimization of such relay assignment and power control has combinatorial aspects [1]. Game theory is used in cognitive radio (CR) networks in which each secondary user (SU) obtains benefits through spectrum sharing by paying prices to primary users (PU) and multi-hop relaying by paying price to nearby SUs. Optimal power allocation methods for SUs are investigated under different assumptions and pricing functions [4]. The performance in cooperative communication depends on careful resource allocation such as relay selection and power control. If the total number of relay nodes increases, the source node can obtain a higher utility value, and the average payment to the relay node shrinks, due to more severe competition among the relay nodes [5].

Two types of resource allocation are commonly used in a wireless network, one is by using CSI and another one is without using CSI[1]. Previously, Game theory was implemented on distributed wireless network to determine the power allocation of the relay nodes for non CSI environment [5]. In this paper, we considered a distributed wireless network where Stackelberg Game theory analyzes the power allocation strategy of the competitive nodes for maximizing the utility function of source and relay nodes. We are applying the Stackelberg gamefor cooperative scenarioto determine the source and relay node utility function. After getting the location of the best relay nodes based on the value of utility function, we are implementing the SPNE considering the environment as non-cooperative. Thus, we get the optimal number of relay nodes and better transmission rate.

2. SYSTEM MODEL AND PROBLEM FORMULATION

In this section, we first describe the system model considered for game formulation. Then we formulate the optimization problem considering a cooperative environment for the selection of optimal number of relay using the Stackelberg game theory. Finally, we will implement the SPNE for Extensive form of non-cooperative environment for getting better transmission rate.

2.1. SYSTEM MODEL

We consider a two-hop AF protocol [1] as our system model and the cooperative transmission consists of two phases. Fig. 1 represents the system diagram, in which there are one source node s, one destination node d, and N relay nodes. In phase 1 source node s broadcasts its information to both destination node d and each relay node. In phase 2, relay node r_i amplifies the signal that it received form source node s and forwards it to destination d with its own transmitted power. Based on the value of price function source node will choose the relay r_i and the relay will act accordingly.

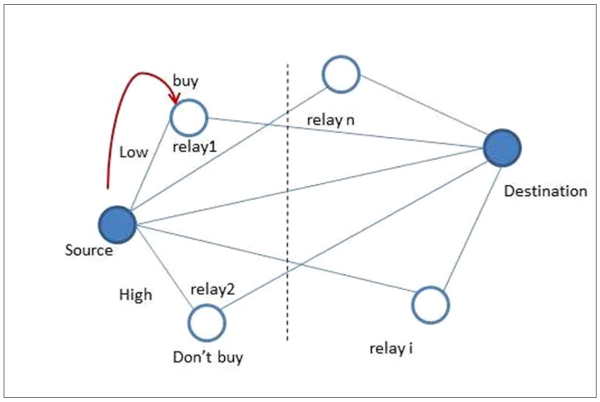


Fig. 1. System Model

In phase 2, multiple nodes transmit the signal towards the destination node d. We consider that, aim of the source node is to maximize its utility function, U_s at least possible payment and the relay node want to earn the payment that covers its forwarding cost and also increases its utility, U_r as many as possible. Cooperative system uses the same total power for all the nodes and bandwidth thus we know the channel bandwidth from local information.

To ensure the equilibrium of the above scenario we are using Stackelberg game theory. Stackelberg Model happens when two or more players compete sequentially on the quantity of output [6]. We deploy Stackelberg game to jointly consider benefits of source node and relay nodes in cooperative communications. In order to attract more buying from the source, the relay adopts a "low-price, high-market" policy to further increase its benefit.

2.2 PROBLEM STATEMENT AND GAME FORMULATION

Referring to the system model in Fig. 1, the received signals at destination node, $z_{s,d}$ and the i-th relay node, $z_{s,ri}$ during Phase 1 can be written as

$$z_{s,d} = \sqrt{P_s G_{s,d}}\, x + n_1 \qquad (1)$$

and,
$$z_{s,ri} = \sqrt{P_s G_{s,ri}}\, x + n_2 \qquad (2)$$

In Phase 2, the received signal at destination node d is

$$z_{ri,d} = \sqrt{P_{ri} G_{ri,\mathrm{d}}}\, x_{ri,\mathrm{d}} + n_3 \qquad (3)$$

where P_s and P_{ri} represents the transmit power at node s and relay node r_i , x is the transmitted signal from node s to node d and node r_i. The average channel gains $G_{s,d}$, $G_{s,ri}$, $G_{ri,\mathrm{d}}$ are modeled as $G_{a,\mathrm{b}} = \mathrm{K}.\mathrm{d}_{ab}^{-q}$ where K is a constant, the distance between the relay node a and b is d_{ab}^{-q} , q is the path loss exponent and n_1, n_2, n_3 are the zero-mean additive white Gaussian noise (AWGN). In general, we can calculate the channel gain based on the bandwidth of the transmission channel, transmit power and noise power. Therefore, a distributed algorithm can be applied for power allocation to the source and relay nodes. In centralized wireless communication system, there must be a node who keeps all the necessary information of the relay nodes called as central node. For maintaining all the nodes information the central node needs more energy, high computation rate. With respect to that, distributed system does not manage all the nodes from a central node. Thus it reduces overhead and computation cost.

Then, the SNR at phase 1 for source to destination direct transmission and at phase 2 for source, i-th relay to destination expressed as [5]

$$\gamma_{s,d} = P_s G_{s,d} / \sigma^2 \qquad (4)$$

$$\gamma_{s,ri,d} = \frac{P_s G_{s,ri} P_{ri} G_{ri,\mathrm{d}}}{\sigma^2 (P_s G_{s,ri} + P_{ri} G_{ri,\mathrm{d}} + \sigma^2)} \qquad (5)$$

Using equation (5), we got the output data rate, $R_{s,r,d}$ for a set of relay node $L = \{r_1,......,r_N\}$, where W is the transmission bandwidth. Based on the behavior of source and relay nodes of a distributed wireless network, we are formulating a game. The achievable rate of the system is defined as

$$R_{s,r,d} = W \log_2 \Big(1 + \frac{\gamma_{s,d} + \sum_{ri \in L} \gamma_{s,r,d}}{\gamma}\Big) \qquad (6)$$

Game theory involves multiple decision-makers and sees participators as competitors (players). In each game players has a sequence of personal moves; at each move, each player have a number of choices from among several possibilities, also possible is the chance or random move. At the end of the game there is some payoff to be gained (cost to be paid) by the players which depends on how the game was played [7]. The actions of several players are interdependent [6]. We are considering the dynamic game: Stackelberg game theory as the each player chooses the strategy to compete sequentially on the quantity of output. A game has incomplete information if one player does not know another player's payoff. Games with complete information in which the players' payoff functions are common knowledge.

In dynamic games of complete and perfect information: first player1 moves, then player2 observes player1's move, then player2 moves according to its strategy and the game ends. According to our system model, all the nodes are players and source *node s* is the leader or the first mover. The game is formed in two steps. Source node moves first and the relay node moves second. Relay nodes choose their strategies only after the strategic choices made by the source node. Optimal power allocation P_{ri} is the payoff between the source and relay nodes. Backward induction is a technique to solve a game of perfect information. According to Backward induction, source node always considers the goal which is last stage of the game, and determines the best move for the player in each case. Considering the goal of the game, it proceeds backwards in time for determining the best move for the respective player, until the goal of the game is reached. Backward induction helps the source node to make plan about the future steps of the relay nodes.

The source node plays the buyer-level game and tries to achieve maximum profit at least possible payoff. The proposed optimization problem for source node s denoted as

$$\max U_s = aR_{s,r,d} - \sum_{ri \in L} \rho_i P_{ri} \tag{7}$$

a, represents the gain per unit of rate and ρ_i expresses the price per unit of power selling from relay node r_i to source node s. U_s is the utility function of the source node and it is concave in nature. Maximized value of the source node utility function U_s depends on the transmission rate $R_{s,r,d}$. According to equation (6), the value of $R_{s,r,d}$ depends on the Signal to Noise Ratio (SNR) at phase1 and phase2 and on the channel gain of source to destination, source to relay nodes and relay nodes to destination node.

Each relay node plays the seller-level game and keep focus to gain not only the forwarding cost, c_i but also achieve as many as extra profit as possible. Utility function of the relay nodes defined as,

$$\max U_r = (\rho_i - c_i) P_{ri} \tag{8}$$

The proposed method is a standard optimization problem where source node aims to minimize the price function by choosing values of variables such that the utility function is maximized and the relay node needs to set the optimal price per unit for the service so as to maximize its own benefit. Price function is used to improve utility that reflects QoS of wireless network. This approach helps the source node to find the relays at relatively better locations.

3. APPLICATION OF GAME THEORY

3.1. OPTIMAL NUMBER OF RELAY SELECTION USING STACKELBERG GAME

In Stackelberg game, equilibrium is reached when a player preemptively expands output and secures larger profits. Source node gets the "first mover advantage". Source node s will gather the information regarding bandwidth and power of the relay nodes. After calculating the price function source node will discard the relay nodes with the bad channel conditions. Then, compute the utility value of source and selected relay nodes. According to Stackelberg game theory the 2nd player is forced to curtail its profit so that the 1st player can get better benefits. Based on the value of the utility function of source node and relay nodes Stackelberg equilibrium is achieved and thus the optimal relays are selected.

We can get the optimal power allocation P_{ri}^* by taking derivative of U_s in equation (7) with respect of P_{ri}

$$P_{ri}^* = \left(\sqrt{\frac{A_i B_i}{\rho_i}}\left(\frac{\sum_{rJ \in L}\sqrt{\rho_J A_J B_J} + \sqrt{\left(\sum_{rJ \in L}\sqrt{\rho_J A_J B_J}\right)^2 + 4\left(1 + \sum_{rJ \in L} A_J\right)W'}}{2\left(1 + \sum_{rJ \in L} A_J\right)}\right)\right) - B_i \tag{9}$$

where, $W' = \dfrac{aW}{\ln 2}$ $A_i = \dfrac{P_s G_{s,ri}}{\sigma^2 + P_s G_{s,d}}$ $B_i = \dfrac{P_s G_{s,ri} + \sigma^2}{G_{ri,\mathrm{d}}}$

By taking derivative of U_r in equation (8) with respect of ρ_i and equating it to zero, we can get the optimal price ρ_i^*

$$\rho_i^* = c_i - \frac{P_{ri}^*}{\dfrac{\delta P_{ri}^*}{\delta \rho_i}} \tag{10}$$

Steps of Relay selection:

Step1: Source node gathers information regarding bandwidth and power for the relay nodes.

Step2: Source node calculates the value of price function of (10) for all the relay nodes and the value must have to be larger than the cost of power for relaying the data. Therefore, there is an optimal price for each relay node based on the position and channel condition of the relay node.

Step3: Calculate the utility function of source (7) and the relay nodes (8).

Step4: Applying the Stackelberg game theory we will find out the optimal number of relays.

3.2. IMPLEMENTATION OF SPNE FOR EXTENSIVE FORM GAME

After selecting the optimal number of relays for cooperative environment using the Stackelberg game theory; we are considering the game as a non-cooperative environment. We are considering an extensive form of game. Thus we solve the game using SubgamePerfectNash Equilibrium (SPNE). In this way, the source node allocates his total power amongst the selected optimal relay nodes and therefore the transmission rate has increased.

In a Nash Equilibrium, each player is doing the best it can, given what its competitors are doing. Nash equilibrium is usually non-cooperative outcomes. Each player chooses the strategy to maximize its profits given its opponent's actions. At the equilibrium, there is no incentive to change strategies, as there is not any chance to improve payoffs.

The natural way to translate a (finite) dynamic interactive decision situation (of complete information) in a game is the extensive-form representation. The extensive-form representation of a game specifies the players (agents) in the game, when each player has the move, what each player can do at each of her opportunities to move, what each player knows at each of her opportunities to move, in games with chance moves: the probabilities assigned to each feasible move, what the outcome is as a function of the actions taken by the players (inclusive the chance playernature) the payoffs of the players (exclusive the chance player) from each possible outcome [11].

One way to get to a prediction of what will (and will not) happen in the extensive-form representation of a dynamic game is first to translate the extensive-form in the associated normal-form and then to apply the concept of NE. There are two problems with this approach. First, dynamic games of complete information typically have many NEs. Secondly (and closely related to the first problem), many NEs in dynamic games involve players choosing non-credible strategies. Subgame Perfect Nash Equilibrium (SPNE) has introduced to solve the problem that rules out non-credible strategies[11].The key feature of a subgame is that, when considered in isolation, it is a game in its own. The concept of a NE can therefore be applied to subgames.SPNE requires a player's strategy to specify optimal actions at every point in the game tree; that is, given that a player finds herself at some point in thetree, her strategy should prescribe play that is optimal from that point on given her opponents' strategies. A subgame of an extensive-form gameG^E is a subset of the game which begins with information set containing a single decision node, contains all decision nodes and terminal nodes that are successors of this node, and it does not cut any information sets.

Therefore the games of the Extensive form can be expressed as G^E

$$G^E = [\, \mathrm{N, H, P, f}, \{\mathrm{I}_i\}_{i\in \mathrm{N}}, \{\mathrm{u}_i\}_{i\in \mathrm{N}} \,]$$

Here, N is the set of players, H represents the set of histories, $P(\Phi)$ and $f(\Phi)$ represents player and nature function, information partition of player i is$I_i$andu_i expresses the payoff of player i.

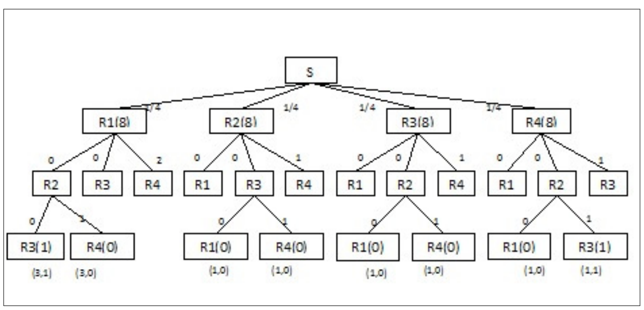


Fig.2. Game tree with chance move considering the no of relay node is 4

Fig. 2 represents a game tree of a dynamic game with perfect information. We consider 4 relay nodes for the simplicity of the representation.We can draw this game tree for any number of relays. Each terminal node lists the players' payoffs arising from the sequence of moves leading to that terminal node. The concept of an information set allows accommodating the possibility. The elements of an information set are a subset of a particular relay's decision nodes.

Step1: Game starts at an initial decision node, source node S makes her move.

Step2: source node s has the element of chance to choose any relay nodes among the 4 relay nodes.(Here, we consider the no of relay node is 4.)The element of chance can be captured in the game tree by including random moves of source node. Source node plays its actions with fixed probabilities According to our game tree the value of nature function is ¼.Each of the 4 possible actions of source node is represented by a branch from the initial decision node. At the end of each branch is another decision node.

Step3: At stage2, 4 decision nodes (Relay node *R1, R2, R3, and R4)* select the next hop based on the value of price function. Each relay node can choose the remaining relay nodes as the next hop. From each decision node there are 3 possible actions which are represented by a branch and ends with a decision node.

Step4: Among the entire decision node at stage3, some of the decision nodes have "zero" power allocation so that these decision nodes cannot take in the next stage of the game. Rest of the decision nodes whose have positive power allocation select the next hop based on the price function.

Step5: At stage 4, only the 2 decision nodes have the positive power allocation. There is an unique connected path of branches from the initial node to each point in the tree. The payoff (power allocation to the relay nodes) leads the source node towards the terminal node.

According to tree shown in Fig. 2 there are two possible subgame; one is for relay node *R1* and another is for relay node *R2.* Payoff of the first subgame, u_1 (1/4,0,0) =3 and payoff of the second subgame, u_2(1/4, 1, 0) =1. Based on the value of player function (price) and payoff function (power) the solution of the Nash Equilibrium is $S > R1 > R2 > R3$

4. Analysis of Simulation Result

The simulation results evaluate the performance of the proposed scheme. Here, the transmit power P_s = 10 mW, bandwidth transmission W = 1 MHz and gain per unit a = 0.85 is considered. In the simulations, the coordinate of source node s and destination node d is (0, 0) and (100, 0) and the complete simulation scenario is depicted in Fig. 3.

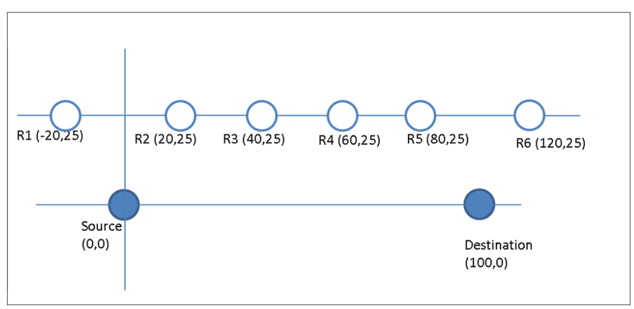


Fig.3. Representation of relay positions as per simulation

Fig. 4(a) shows the transmission rate according to the number of relay nodes. It shows that the transmission rate is increased with the number of relays when the number of relays is less than 4.The relay management is complex for the higher number of relays. Moreover, Transmission rate has risen negligibly for the higher number of relays (between 4th and 6th relay nodes).Therefore, selecting too many relays may not be practically useful for the system. In comparison of two protocols, for non-cooperative relays transmission rate is higher than cooperative relays for the first two relays. From the 3rd relay it is lower than the cooperative transmission rate.

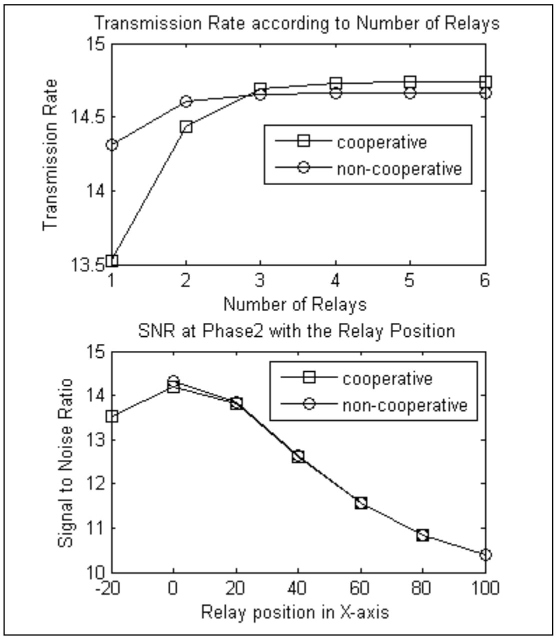


Fig. 4(a) Transmission rate and the number of relays (b) Signal to Noise Ratio and the Relay position in X-axis

Fig. 4(b) depicts the relation between the SNR and the relay position where both cooperative and non-cooperative schemes achieve almost similar performance.We also found from Fig. 4(a) and 4(b) that, in this simulation scenario, the near optimal number of relay nodes between source and destination is 4 and if the position of the relay node *r* is far away from the source node *s*, the SNR is getting lower.

For both cooperative and non-cooperative environment, we get the highest SNR values when the relay locates between (0, 25) and (40, 25) in X-axis. Thus it proves that, for better transmission rate we have to choose the relay nodes that located within the source and destination nodes and relatively closer to the source node.

Fig. 5 represents the relation of power allocation and source node utility function with respect to relay position in X-axis. In 5(a) power allocation is higher for the relay nodes that are closer to the source node. The allocation of power from source to relay nodes is proportionally decreasing towards the destination. At non-cooperative environment, source node allocates relatively more power for the first two relay node based on the value of the price function. From fig. 5(b), we got that the source utility is maximum at (20, 25) for both cooperative and non-cooperative scenario. Source node utility value increases sharply till to the relay node (20, 25) and after that it decreases linearly.

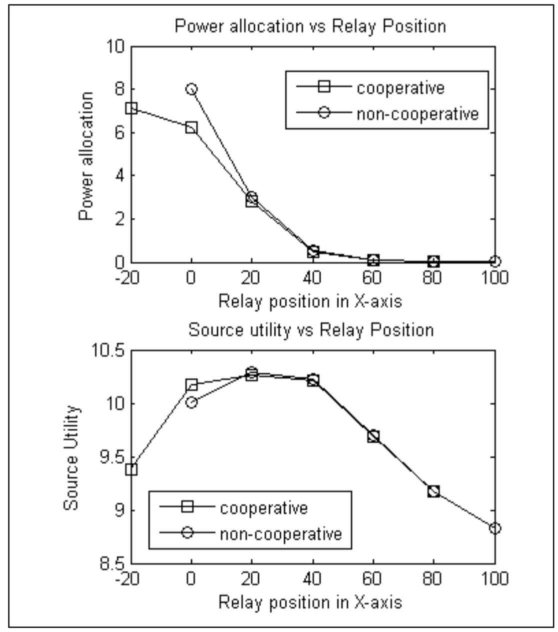


Fig 5 (a) Power allocation vs Relay position in X-axis (b) Source Utility vs Relay position in X-Axis

For comparing the performance of the proposed game with the centralized scheme, we set up the simulation as follows: there are two relay nodes and one (*s*,*d*) pair. One relay node is fixed at coordinate (60, 25) and other node is fixed between (20,25) and (80,25).

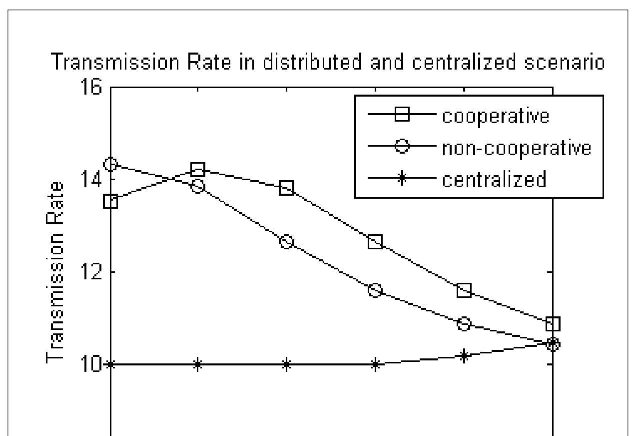


Fig. 6 Transmission Rate in distributed and centralized environment

Fig. 6 describes the comparison in transmission rate for distributed and centralized relay selection. The transmission rate of the distributed (cooperative and non-cooperative) relay selection is higher because the power allocation is maximum [14]. In our proposed method, the source node calculates the power of relay nodes and makes its decision based on that. The transmission rate decreases gradually in distributed communication and slightly increased for centralized communication when relay nodes are getting away from the source node. The transmission rate is nearly equal for both distributed and centralized method when the number of relays is 6.

5. CONCLUSION

Previously, game theory was used to determine the power allocation between relay nodes in different scenario. In this paper, first we implement the Stackelberggamefor cooperative environment to find out the optimal number of relay based on the value of the source and relay node utility function.The relays which stand closer between source node and destination node have higher transmission rate and utility value. Simulation results showed that we getbetter transmission rate if the number of relay node is not more than 4. Finally, we have considered a non-cooperative distributed wireless environment and impose SPNE to improve the transmission rate.The proposed non-cooperative game also showed similar result as cooperative game. For large number of relay nodes SPNE does not provide optimal solution. By selecting the optimal number of relays we can reduce the complexity of relay management. The location range of the selected relays and the number of optimal relay nodes is dependenton the simulation environment. The result may also vary with respect of the simulation parameters. But the proposed method is global and can be implemented in any practical relay network.

REFERENCES

[1] Naeem, M., A. Anpalagan, et al. "Resource Allocation Techniques in Cooperative Cognitive Radio Networks." Communications Surveys and Tutorials, IEEE communications surveys & tutorials, vol. 16, no. 2, 2nd quarter 2014.

[2] KanchanVardhe, Daryl Reynolds and Brian D. Woerner, “Joint Power Allocation and Relay Selection for Multiuser Cooperative Communication", IEEE transactions on wireless communications, vol. 9, no. 4, April 2010.

[3] Kais Ben Fredj, Sonia Aissa, “Performance of Amplify-and-Forward Systems with Partial Relay Selection under Spectrum-Sharing Constraints”, IEEE transactions on wireless communications, VOL. 11, no. 2, February 2012.

[4] Yong Xiao, Guoan Bi, DusitNiyato, “Game Theoretic Analysis for Spectrum Sharing with Multi-Hop Relaying” , IEEE transactions on wireless communications, vol. 10, no. 5, May 2011.

[5] B. Wang, Z. Han and K.J.R. Liu, “Distributed Relay Selection and Power Control for Multiuser Cooperative Communication Networks Using Stackelberg Game”, IEEE Trans. Mobile Computing, vol.8, no.7, July 2009.

[6] Theodore L. Turocy, Bernhard von Stengel, “Game Theory”, CDAM Research Report, LSE-CDAM-2001-09.

[7] L. Pavel, “Game Theory for Control of Optical Networks, Static & Dynamic Game Theory: Foundations & Applications”, DOI 10.1007/978-0-8176-8322-1_2, Springer Science Business Media, New York 2012.

[8] Alsharoa, A., H. Ghazzai, et al., “Near-optimal power allocation with PSO algorithm for MIMO cognitive networks using multiple AF two-way relays”, IEEE ICC, 2014.

[9] C. Chang, P. Lin and S. Su, “A low-interference relay selection for decode-and-forward cooperative network in underlay cognitive radio”, IEEE CROWNCOM, June 2011.

[10] Yuanyuan Zhang, Yi Ma and Rahim Tafazolli, " Power Allocation for Bidirectional AF Relaying over Rayleigh Fading Channels", IEEE communications letters, vol. 14, no. 2, February 2010.

[11] Rudolf Kerschbamer, "Commitment and Information in Games", National Centre for Econometric Research: NCER, Annual Report, 2009.

[12] Tamer Basar, "Lecture Notes on Non-Cooperative Game Theory", July26,2010

[13] Guillermo Ordoñez, "Notes on Nash Equilibrium -. ECON 201B # Game Theory." January19,2006

[14] L. Giupponi and C. lbars, “Bayesian Potential Games to Model Cooperation for Cognitive Radios with Incomplete Information,” Proc, IEEE ICC, June 2009.

Implementation of Improved Harmony Search Based Clustering Algorithm in Wireless Sensor Networks

Anamika Dey[1]
Department of CSE
University of Asia Pacific
Dhaka, Bangladesh.
Email: anamika.dey.bd@ieee.org

Tamal Sarkar[2]
Department of EEE
Bangladesh University of Engineering & Technology, Dhaka.
Email: tamal.sarkar.bd@ieee.org

Md. Arif Ullah[3], Nasrin Nahar[4]
Department of CSE
University of Asia Pacific
Dhaka, Bangladesh.

***Abstract*- The essential restriction in wireless sensor network (WSN) is constrained energy supply. The center of this paper is to put accentuation on building up a brought together, energy proficient clustering algorithm in light of Improved Harmony Search (IHS) to augment lifetime of sensor network and assessment of simulation situations its execution on a few WSN. The entire wireless sensor network is parceled into a few bunches by IHS algorithm. For streamlining, energy utilization of the nodes in WSN is the essential thought of actualizing the algorithm. The assessment of the IHS clustering algorithm against its counterparts was done on the premise of network lifetime and the measure of information sent to the Base Station (BS). Later on, the earlier practiced algorithms, for example, LEACH, LEACH-C and Cuckoo Search protocols were tried against the advocated protocol.**

***Keywords*- Improved Harmony Search (IHS), Wireless Sensor Network (WSN), Cuckoo Search (CS), Base Station (BS), Cluster Head (CH), Pitch Adjustment Rate (PAR), Bandwidth (bw).**

I. INTRODUCTION

With the expansion of automation system, smart grids, security surveillance systems and other networking and cyber-security issues, wireless sensor network (WSN) has been an emerging and ultimate solution for efficient and reliable data acquisition and communication medium. [12-15,17] Essentially, a WSN comprises of a huge amount of sensor nodes which are little power consuming and miniature in terms of size. [1] Nodes are well capable of working as autonomous devices and can be deployed in different types of environments. Still pragmatic application of WSN is posing some challenges that need to be met. The primary urging issue is life-time extension of the deployed sensor nodes. The sensor nodes are furnished with some measure of detecting, control, information preparing, and imparting parts. [4] WSNs are critically resource constrained by memory, limited power supply, and communication bandwidth and processing performance. [4,20-22]Accordingly in verging on each situation, sensor nodes depend on a restricted supply of energy (utilizing batteries). Supplanting these energy sources in the field is normally not achievable, and in the meantime, a WSN must capacity at any rate for a given time or as far as might be feasible. Thus, most existing works (e.g. bunching, drawing out lifetime) in WSNs range are managing energy effectiveness. [16]

For highly energy efficient operation of WSNs, Cluster-based routing protocols are considered as well-established techniques. They also have special advantages related to effective communication and scalability [6]. In recent days, numerous cluster-oriented protocols have been offered with a view to extending the network period. Low Energy Adaptive Clustering Hierarchy (LEACH) [3] is an emblematic cluster-based protocol that uses a distributed clustering development algorithm. Here, the choice of CHs is constructed on a predefined possibility; other nodes pick the cluster to join by approximating the adjacent distance to the CHs elected. Nevertheless, LEACH does not initiate the identical circulation of CHs in the network. Moreover, some nodes are selected as CHs for another rounds of process, thus consume extra energy than the others. In [4], LEACH-centralized (LEACH-C) is advocated as an upgrading of LEACH which uses a centralized clustering procedure to build the clusters. LEACH-C augments the system performance by producing improved clusters and by scattering the cluster head nodes all through the network. Yet, both LEACH and LEACH-C do not discover the optimal choice of cluster based on any objective function. To have control over the network energy costs proficiently, clustering is a tactic used widely. It abates the number of sensor nodes that include in long distance communication with base station (BS) and allocates the energy spending evenly amongst nodes in the sensor network.

This paper describes about the advancement of low energy clustering method based on Improved Harmony Search (IHS) motivated to prolong the lifetime of sensor network and estimate the performance on a number of WSN over simulation environments. The protocol divides the entire network into some clusters by IHS algorithm. The IHS algorithm incorporates HS algorithm with the fine tuning property of mathematical methods and can outdo either one exclusively. The implementation of IHS algorithm in wireless sensor networks is the novel feature of the research work. The algorithm deliberates energy consumption of the nodes in the optimization procedure.The established protocol later was verified against the earlier established protocol like LEACH,

LEACH-C and CS which are the recently used bio inspired optimization algorithms.

II. THE SYSTEM MODEL

A. Network Structure

We assume a network structure comparable to those utilized in [9] and [11], with the accompanying characteristics: the BS is unaltered and situated in the sensor network area. All sensors are fixed and perform detecting. All sensor nodes perform their task so often and send the data/information to the BS. Every hub is fit for working in cluster head (CH) mode and detecting mode relying upon the level of residual energy. Information aggregation is utilized to diminish the total message sent.

B. Wireless Energy Model

The wireless energy model for the sensors utilized as a part of our protocol depends on the first order wireless model as utilized as a part of [10,18]. In this model, to accomplish an agreeable Ratio of Signal-to-Noise (SNR) in transmitting an l-bit data over a separation d, the transmitting energy is given by:

$$\begin{aligned} E_{TX}(l,d) &= l \cdot E_{elec} + l \cdot \varepsilon_{FS}\, d^2, \quad \text{if } d < d_0 \\ &= l \cdot E_{elec} + l \cdot \varepsilon_{TR}\, d^4, \quad \text{if } d \geq d_0 \end{aligned} \tag{1}$$

Where, E_{elec} is the degenerated energy for each bit of information which is utilized to run the transmitter or the receiver. ε_{FS} and ε_{MP} shifts as for transmitter amplifier model that is utilized. d_0 indicates the threshold estimation of the transmission distance. Thus, to get l bit data packet, the energy consumed by the radio is:

$$E_{RX}(l) = l \cdot E_{elec} \tag{2}$$

III. DETAILING OF IMPROVED HARMONY SEARCH ALGORITHM

Due to the strict energy constraint that the sensor nodes, deployed in real environment and powered by battery, may not have energy replenishment capabilities, it is important to balance the energy consumption of network for the IHS algorithm, so as to prolong the network lifetime as far as possible. In this section, we propose a new energy-efficient algorithm for WSNs, which is based on IHS algorithm.

Firstly, a basic implementation of the routing process of proposed algorithm for WSNs is presented. Next, the Energy Efficient Improved Harmony Search Based algorithm is explained. Finally, the data transmission and simulation is presented using this IHS algorithm. The main issue of the entire simulation and resulting part is the energy efficiency.

A. Improved Harmony Search Algorithm

Harmony Search (HS) algorithm was as of late created in a correspondence with music spontaneous creation process, where music players improvise the pitches of their instruments to get better harmony [5].
Improved harmony search algorithm procedure is segmented below:

- Set the optimization issue and algorithm parameters.
- Set up the harmony memory.
- Make up a new harmony.
- Update the harmony memory.
- Checking bring to a halt criterion.

These steps are depicted in the next five subdivisions.

Step 1: Set the optimization issue and algorithm parameters*:*

Initially, it is needed to minimize the distance of intra cluster and advance the energy utilization of the network. The improvement issue is determined as takes after:

Minimize *f(x)* subject to $x_i \in X_i = 1,2,\ldots\ldots, N$ (3)

Where f(x) is a goal capacity; x is the arrangement of every choice variable Xi ; N is the quantity of choice variables, X_i is the arrangement of the conceivable scope of qualities for every choice variable, that is $U_{xi} \geq Xi \geq L_{xi}$ and U_{xi} and L_{xi} are the upper and lower limits for every choice variable. The HS algorithm parameters are likewise determined in this stride. The size of harmony memory (HMS), or the quantity of arrangement vectors in the harmony memory; harmony memory considering rate (HMCR),the number of improvisations(NI), pitch adjusting rate (PAR). The harmony memory (HM) is a memory area where all the arrangement vectors (sets of choice variables) are put away. This HM is like the hereditary pool in the GA [2,21]. Here, PAR and HMCR are the parameters those are utilized to enhance the vector arrangement.

Step 2: Set up the harmony memory

In Step 2, the matrix is loaded called Harmony Memory (HM) with the same number of arbitrarily produced vectors solution as the HMS. Every row of the HM is an arbitrary solution for the optimization issue and after that the estimation of the target capacity is figured for every harmony vector.

$$\text{HM} = \begin{bmatrix} x_1^1 & x_2^1 & \cdots & x_{N-1}^1 & x_N^1 \\ x_1^2 & x_2^2 & \cdots & x_{N-1}^2 & x_N^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_1^{\text{HMS}-1} & x_2^{\text{HMS}-1} & \cdots & x_{N-1}^{\text{HMS}-1} & x_N^{\text{HMS}-1} \\ x_1^{\text{HMS}} & x_2^{\text{HMS}} & \cdots & x_{N-1}^{\text{HMS}} & x_N^{\text{HMS}} \end{bmatrix}. \tag{4}$$

Step 3: Make up a new harmony

Subsequent to characterizing the HM as appeared in Step 2, the act of spontaneity of the HM is finished by producing another harmony vector [x_1, x_2,…, x_p]. Each segment of the new harmony vector X_j is produced utilizing (5) based upon the HMCR characterized in Step 1. The HMCR changes from 0 to 1, is the rate of picking one value from the historical values put away in the HM, while (1-HMCR) is the rate of arbitrarily selecting one value from the conceivable scope of values.

$$x_i' \leftarrow \begin{cases} x_i' \in \{x_i^1, x_i^2, \dots\dots, x_i^{HMS}\} & with\ probability\ HMCR \\ x_i' \in X_i & with\ probability\ (1-HMCR) \end{cases} \quad (5)$$

Each part got the consideration by the memory is inspected to figure out if it ought to be pitch-adjusted. This operation utilizes the PAR parameter that is the rate of pitch which takes after:

$$\text{Pitch adjusting decision for } x_i' \leftarrow \begin{cases} Yes\ \ with\ probability\ PAR \\ No\ with\ probability\ (1-PAR) \end{cases} \quad (6)$$

The estimation of (1 - PAR) sets the rate of doing nothing.

The conventional HS algorithm utilizes settled worth for both bw (Bandwidth) and PAR. Here, HS technique PAR and bw values balanced in first step (Step 1) and can't be changed amid new generation. The principle downside of this technique shows up in the number of iterations the algorithm needs to locate an ideal solution.

The major difference between IHS and conventional HS method is in the process of selecting PAR and bw. To get a developed performance from the HS algorithm and sort out the drawbacks for constant values of PAR and bw, IHS algorithm utilizes variable PAR and bw in creating step (Step3). PAR and bw get changed dynamically with generation number. [12,19]

Step 4: *Update harmony memory*

If the new harmony vector, $x' = \{x_1, x_2, \dots, x_p\}$ is superior to the most noticeably bad harmony in the HM, judged regarding the target capacity esteem, the new harmony is incorporated into the HM and the current most exceedingly terrible harmony is barred from the Harmony Memory.[12]

Step 5: *Checking bring to a halt criterion*

In the event that the halting foundation (most extreme number of act of spontaneity) is fulfilled, calculation is terminated. Something else, Steps 3 and 4 are repeated.

.

IV. *Cluster Setup Using IHS Algorithm*

The IHS algorithm is a brought together clustering algorithm where the BS registers and divides the nodes into many clusters based on positional data and energy level of the nodes. Let us assume a structure of sensor nodes of N numbers and that nodes are arranged in k clusters: C_1, C_2,..., C_p. The data sending process is divided into rounds. In every round, the CHs gather information from every single group part and transport to the destination (BS). The entire execution incorporates 2 stages: one clustering model and another is data transmission.

A. Process of Clustering model

In this clustering framework fragment, these sensor nodes transmit information to the BS with the data of their physical position and level of energy. Taking into account this information to the BS will process the normal energy level of every single dynamic node. Just the nodes which have remaining energy upper than the normal level are qualified for being an immediate cluster head for the running round or circle. To improve the determination of cluster heads, IHS Algorithm is raced to discover optimal p cluster heads, the target function in [6] is utilized to calculate the minimum, as appeared underneath:

$$f_1 = max_p \ \left\{\sum_{\forall i \in Cp} \ d\left(n_i, CH_p\right)/\left|C_p\right|\right\} \quad (7)$$

$$f_2 = \sum_{i=1}^{N} E\ (n_i)\ /\sum_{p=1}^{K} E\ \left(CH_p\right) \quad (8)$$

$$f(x) = \alpha \times f_1 + (1-\alpha) \times f_2 \quad (9)$$

In this goal function, *f1* is the most extreme of the Euclidean space among nodes, node i∀i ∈ cluster Cp, to their cluster heads CH_p and |Cp| is the quantity of nodes that fit in with cluster Cp. In the meantime f2 is the proportion of introductory energy of every active node, E(nodei) in this network structure with the aggregate existing drive of the running cluster head, E(CHp), in the running round. The steady α demonstrates the commitment of f1 and f2 in the target capacity f(x). This target function has a tendency to diminish the intra-cluster average distance among sensor nodes and their cluster head and in addition to upgrade the energy proficiency of the structure of the network. Once the cluster is framed, the non-cluster head nodes transmit information towards the BS through the cluster heads. The procedure of selecting clusters is rehashed each cycle of exchanging information among sensor nodes. Amid the transmission from the nodes to CH, remaining energy of every node are joined to the information packet, this data helps the BS to choose the most appropriate cluster head and cluster at the following round. Taking into account the number of the dynamic nodes inside of the cluster, the CH makes a timetable in view of Time Division Multiple Access (TDMA).

B. Data Transmission:

Initially, the CHs are allocated. The schedule of the process of transmission is prepared, the nodes of the sensor start to transmit data to the CHs. The non-cluster head nodes send their data to the CH. The CH accumulates the total data of a cluster and sends the total data to the BS. The total data sending process from a CH works as a cycle after completing one cycle in new CH is selected based on highest node energy. Moreover, TDMA convention is utilized; non-cluster head nodes switch their transmitting node amid the process and kill subsequent to completing transmission. Information amassing and combination is done at the CHs, along these lines the measure of data is diminished, and the CHs just lead the compacted information to the BS.

C. Flow Chart of Cluster Head Selection Using IHS Algorithm:

In the figure 1, the flow chart of cluster head selection in Wireless Sensor Network is illustrated below:

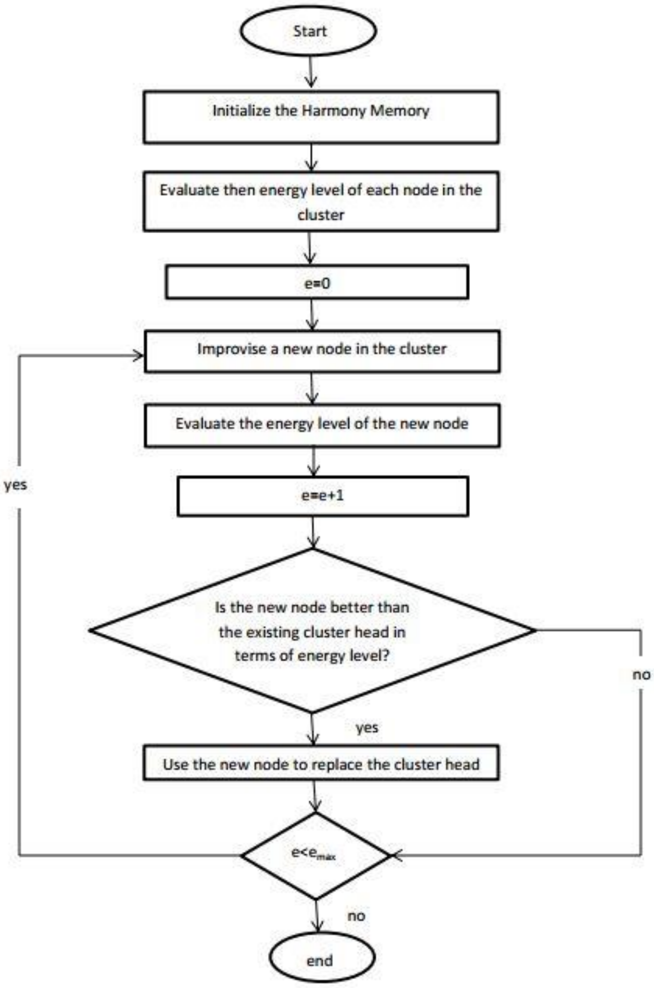


Fig 1: Flow Chart of CH selection using IHS algorithm

V. *Simulation Result and Discussion*

Fig 2 and 3 respectively delineates the life-span after the demise of total nodes and the total network life-span, characterized by measure of nodes dynamic over cycle (100 nodes, 1000 rounds). It demonstrates that the IHS convention surpasses the aggregated network life time of LEACH, LEACH-C, and CS, by about 54 %, 38 %, and 8 % correspondingly. This change depends on two reasons. Firstly, with least intra cluster distance and optimal CH dispersion over the network, the IHS convention indicates enhanced system apportioning. Thus, the energy scattered by all nodes for correspondence is lessened. Furthermore, the IHS system received in the convention creates an arrangement of good exchange offs where the estimations of the cost capacity are middle of the road to the network prerequisites.

Fig 4 demonstrates the aggregate sum of information messages got at BS by every one of the conventions and algorithms. Improved Harmony Search (IHS) enhances the information conveyance by elements of 11% over CS, 54% over LEACH-C and 161% over LEACH. The thought behind this is, our protocol can take the advantage of selecting the high energy hub as a CH by mulling over the leftover energy of the CH hopefuls furthermore the base distance between the nodes and their CHs by actualizing advanced cost capacities. Subsequently, more information messages are conveyed to the BS.

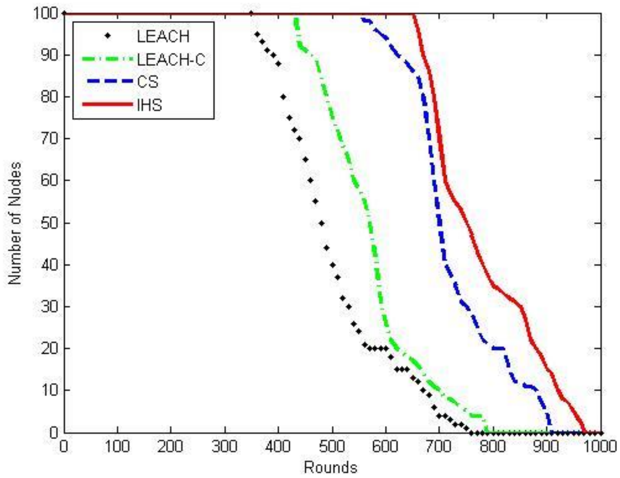


Fig 2: Network lifetime for the systems: Leach, Leach-C, CS, IHS

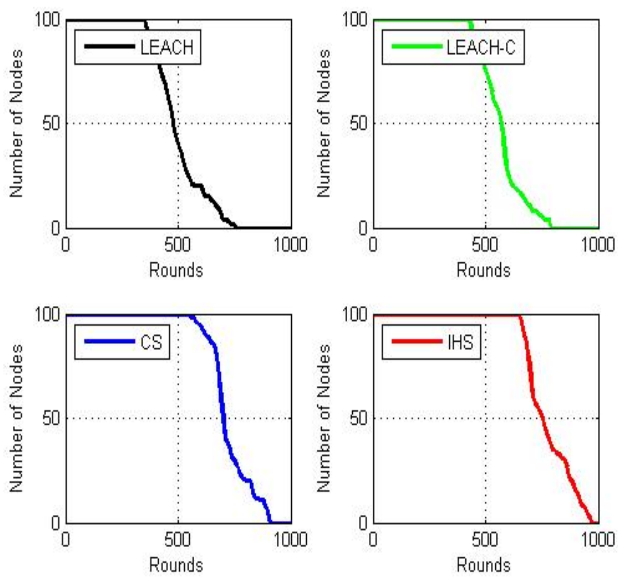


Fig 3: Network lifetime for individual protocols

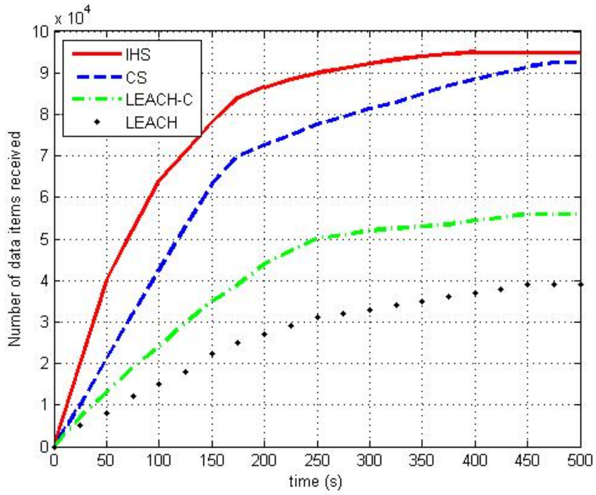


Fig 4: Entire data received at BS with respect to time

VI. *Conclusion and Future work*

This paper delineates a various leveled clustering calculation with energy-awareness for WSNs utilizing the IHS calculation. With a specific end goal to minimalize the intra-cluster distance along with improving the energy utilization of the network, the IHS algorithm is utilized to make cluster structure. The cost function was characterized in a style that takes into count the most elevated distance in the midst of the non-CH node along with its related CH, and the leftover force of CH hopefuls in CH choice algorithm. Reproduction results delineate that the proposed IHS based convention gives upgraded network life span and shows capacity of conveying more information to the BS contrasted with LEACH, LEACH-C and CS based clustering methods. Furthermore, the proposed protocol produces better cluster development by just as allotting the CHs throughout the network zone. The upcoming attempt is to extend the work to the cross-layer optimizing in the middle of inquiry and directing techniques. Also, it can be stretched out by considering multi-hop communication as a part of CHs to build energy efficiency. Besides, hybridization of a few other learning methods should be possible to decide energy efficient cluster structures.

VII. *References*

[1] I. F., Akyildiz, W. L., Su, S., Yogesh, C., Erdal, "A survey on sensor networks" *IEEE Commun. Mag.*, vol. 40, no. 8, pp. 102–114, 2002.

[2] Z.W., Geem, J.H. ,Kim, G.V.,Loganathan, "Harmony search optimization" application to pipe network design, Int. J. Model. Simul. 22 (2) (2002) 125–133.

[3] W., Heinzelman, A., Chandrakasan, H., Balakrishnan, "Energy efficient communication protocols for wireless sensor networks" In Proceedings of the 33rd Annual Hawaii International Conference on System Sciences (HICSS), Hawaii, USA, January 2000.

[4] I.F., Akyildiz, W., Su, Y., Sankarasubramaniam, E., Cayirci, "Wireless sensor networks"a survey. Computer Networks.38, 393–422 (2002).

[5] K.S., Lee, Z.W., Geem, " A new meta-heuristic algorithm for continues engineering optimization: harmony search theory and practice" Compute. Meth. Appl. Mech. Eng. 194 (2004) 3902–3933.

[6] J. N. ,Al-Karaki ,A. E., Kamal, "Routing techniques in wireless sensor networks" A survey, *IEEE Wireless Commun. Mag.*, vol. 11, no. 6, pp. 6–28, Dec. 2004.

[7] P., Honeine, F., Mourad, M., Kallas, H., Snoussi, H.,Amoud, C., Francis,: "Wireless sensor networks in biomedical: Body area networks" IEEE (2011).

[8] Z., Zhang, " Investigation of wireless sensor networks for precision agriculture" American Society of Agricultural and Biological Engineers. pp. 1157–1164. American Society of Agricultural and Biological Engineers (2004).

[9] W.R., Heinzelman, A., Chandrakasan, H., Balakrishnan, "Energy-Efficient Communication Protocol for Wireless Microsensor Networks" 33rd Hawaii International Conference on System Sciences. pp. 1–10. IEEE, Hawaii (2000).

[10] W.B., Heinzelman, A.P., Chandrakasan, H., Balakrishnan,: "An application-specific protocol architecture for wireless microsensor networks" IEEE Transactions on Wireless Communications. 1, 660–670 (2002)

[11] M.,Mahdhavi, M.,Fesanghary, E.,Damangir, "An Improved Harmony Search Algorithm for Solving Optimization Problems" Elsevier Journal of Applied Mathematics and Computation 188 (2007) 1567–15

[12] Florence, M.L., Swamydoss, D.: A SURVEY ON WIRELESS SENSOR NETWORK ARCHITECTURE , PROTOCOLS AND APPLICATIONS. Journal of Global Research in Computer Science. 2, (2011).

[13] Honeine, P., Mourad, F., Kallas, M., Snoussi, H., Amoud, H., Francis, C.: Wireless sensor networks in biomedical: Body area networks. IEEE (2011).

[14] Bahrepour, M. et al. (2010). Distributed Event Detection in Wireless Sensor Networks for Disaster Management. In *2010 International Conference on Intelligent Networking and Collaborative Systems*. 2010 IEEE. pp. 507–512

[15] Yang, X.-S. Koziel, S. (2011). Optimization Algorithms. In *Computational Optimization, Methods and Algorithms*. pp. 13–31.

[16] P., Honeine, F., Mourad, M., Kallas, H., Snoussi, H.,Amoud, C., Francis,: "Wireless sensor networks in biomedical: Body area networks" IEEE (2011).

[17] K. Yang, Wireless Sensor Networks, John Wiley & Sons, 2014.

[18] E. Amiri, H. Keshavarz, M. Alizadeh, M. Zamani, T. Khodadadi, Energy efficient routing in wireless sensor networks based on fuzzy ant colony optimization, International Journal of Distributed Sensor Networks 2014 (2014)1–17.

[19] N. Harris, G. Merrett, H. Zia, Catchment-scale water quality monitoring,control and management framework using collaborative wireless sensor networks, in: 11th International Conference on Hydroinformatics, New York,2014.

[20] G. Hackmann, W. Guo, G. Yan, Z. Sun, C. Lu, S. Dyke, Cyber-physical codesign of distributed structural health monitoring with wireless sensor networks, Parallel and Distributed Systems, IEEE Transactions on 25 (1) (2014) 63–72.

[21] S. A. Haque, M. Rahman, S. M. Aziz, Sensor anomaly detection in wireless sensor networks for healthcare, Sensors 15 (4) (2015) 8764–8786.

[22] P. Kuila, P. K. Jana, A novel differential evolution based clustering algorithm for wireless sensor networks, Applied Soft Computing (2014) 414–425.

[23] D. Sahin, V. C. Gungor, T. Kocak, G. Tuna, Quality-of-service differentia-tion in single-path and multi-path routing for wireless sensor network-based smart grid applications, Ad Hoc Networks 22 (2014)

[24] N. Harris, G. Merrett, H. Zia, Catchment-scale water quality monitoring, control and management framework using collaborative wireless sensor networks, in: 11th International Conference on Hydroinformatics, New York,2014.

Optimal Power Allocation for Multichannel Cognitive Radio Systems Using Stackelberg Game

Mohammad Obaidur Rahman and Asaduzzaman
Department of Computer Science and Engineering
Chittagong University of Engineering and Technology, Chittagong-4349, Bangladesh
Email: obaidur_91@cuet.ac.bd and asad@cuet.ac.bd

Abstract –We consider a non-orthogonal multiple access scheme e.g. CDMA based cellular multi-channel cognitive radio network (CRN). Here multiple secondary users share a common set of channels simultaneously. For such an environment, we develop a distributed power allocation approach based on a game theoretic formulation among secondary users to satisfy their rate requirement. We model the problem using a two-stage leader-follower game known as Stackelberg game where base station and users are considered as leader and followers respectively. Since the spectrum is licensed to a primary network, base stations of the cellular secondary network should protect primary users against interference caused by secondary communications. We are interested for downlink power allocation with pricing that maximizes utilities of both BS and CRs in presence of primary network. A nonlinear relationship which is developed between bit rate and SINR is solved by game theory. Simulation results show that our proposed power control scheme provides higher SINR levels (or equivalently higher data rates) while consuming lower average transmit power of secondary users.

Keywords–Cognitive radio network(CRN), power allocation, non-cooperative game, Stackelberg game, Nash equilibrium(NE), pricing, QAM, ASK, PSK and BER.

I. INTRODUCTION

Power control of users of cognitive radio is a critical issue. Spectral efficiency and Energy efficiency are also interdependent with power control. Decreasing power consumptions in devices of cognitive radios affects quality of services e.g. bit error rate. So adaptive power control scheme is required. Earlier solutions are Calculus based approach, Convex Optimization, Karush-Khun-Tucker solution, Variational-inequality etc[1]. Each of the method have many limitations. Such as they are (a) time consuming (b) large solution space and (c) failure to solve nonlinear constraints. Game theory can substantially solve the above problems. It can also converge the solution in short time. Game has generally two forms non-cooperative and cooperative.

In this paper we apply Stackelberg non-cooperative game for power control of cellular cognitive radio network. Players of Stackelberg game classified as leaders and followers. Here a base station and secondary cognitive radios in each cell is considered as a leader and followers respectively. Leader(BS) move first and followers are secondary cognitive radios(CRs)[2]. Consequently each of BSs have more benefit than follower. This paradigm best suits for cells of cognitive radio networks. We are interested in developing a decentralized scheme for downlink power allocation with pricing that maximizes utilities of both BS and CRs in presence of primary network. Each BS wants to earn higher income from the CRs in its cell while spending less transmit power. Each CR is interested in gaining higher level of SINR (or equivalently data rate) with high QoS(e.g. bit error rate) while paying less to the associated BS. With the objective of maximizing utilities and upper bound of the power of secondary cognitive radios are such that upper bound of bit error rate(BER) is maintained and primary user can tolerate interference limit.

II. EXSISTING WORK

Nowadays application of Game theory in wireless communication is one of the substantial matter. Previous method of calculus is restricted in finding critical point by avoiding point of inflection. Convex Optimization, Karush-Khun-Tucker solution methods are applied to reduce search space and to find optimal solution avoiding suboptimal solutions[1]. The above methods fail if any nonlinearity exists between two parameters.

Power allocation of cognitive radio network has generally two forms: Centralized architecture and game theory based distributed. A centralized architecture which requires more control signals and parameters can difficult to control and computationally hard . A user based distributed architecture easy to control but sometimes gives suboptimal result[3]. However, the latter more preferable as it is computationally easier and requires less power thus also energy efficient.

Power of spectrum sensing and power allocation of nodes are considered now simultaneously. In [4], the authors formulated as a distributed non-cooperative power allocation game to maximize the total throughput of the cognitive radio users thus optimizing jointly both the detection operation and the power allocation. They took into account the influence of the sensing accuracy. In [5] the authors proposed an algorithm based on the VCG (Vickrey–Clarke–Groves) model in a non cooperative game for spectrum allocation. They done this between secondary transmissions that guarantee a required minimum data rate for both PUs and SUs, assuming a fixed value of the bit error rate. In[6], the authors considerd a sensing-based spectrum sharing scenario in cognitive radio networks where the overall objective is to maximize the sum-rate of each cognitive radio user by optimizing jointly both the detection operation based on sensing and the power allocation.They took account the influence of the sensing accuracy and the interference limitation to the primary users. The resulting optimization problem for each cognitive user is non-convex, thus leading to a non-convex game.Thus they presented a new challenge when analyzing the equilibria of this game where each cognitive user represents a player.

At present Stackelberg non-cooperative game is widely used for resource allocation of cognitive radio network. Stackelberg game has three forms. They are as follows: (a) multiple leaders and multiple followers, (b) single leader and multiple followers and (c) single leader and single follower. Our problem is best suitable with second form[2]. Because in each cell of cognitive radio network there is one base station(BS) and multiple cognitive radios(CRs).

In [7] the authors addressed the application of game theory to resource management of a cognitive network in the downlink side. As the optimum resource allocation problem is NP-hard, they used game theory to propose a distributed scheme for power allocation to guarantee communications of primary networks and provide required signal-to-noise plus interference of cognitive radios. In return, each base station charges the cognitive radios in its cell for their provided received SINR. They modelled the problem using a two-stage Stackelberg game.

From engineering point of views cellular type network is feasible. It has less information overhead[8]. Additional information theoretic facilities is added to such a network. We consider multiuser with multichannel in each cell of cognitive radio network .Co-exist of multipleuser in a channel increases spectral efficiency. Power control should meet the following requirements: (a) Secondary users rate requirements : Power consumption should such that users' total rate be greater than a specified limit thus poorer channels get benefit. (b) Channels' Rate requirement: Power consumption should such that channels' total rate be less than a specified limit. (c) Channels' power requirement: Power consumption should such that channels' total power should be less than a specified limit such that primary users can tolerate interference[9]. All these requirements are often contradictory with each other.

We solve this problem by applying Stackelbergh game and gets equilibrium named as Stackelbergh equilibrium(SE)[11]. If a single user deviate from the equilibrium but utility does not change then the equilibrium is called nash equilibrium(NE). If single point satisfy both the equilibrium conditions then power allocation among the users minimized. We also consider bit error rate constraints for three modulation schemes. The modulation schemes are as follows: quadrature amplitude modulation (QAM), amplitude shift keying (ASK) and Phase shift keying (PSK) scheme with an adaptable modulation order M. Thus quality of service is assured. This is our main goal. In previous works either QAM or PSK modulation scheme is studied in [9] and [12].

So, in this paper our contributions is as follows:

- Minimize Problem search space, Satisfy QoS (BER) and solve linear and nonlinear relation by Game theory for three modulation scheme QAM, ASK, PSK and compare their performances.
- Gives user based distributed solutions thus computational complexity reduces.

III. GAME AND SYSTEM MODEL

A. STACKELBERG GAME

There are two types of players in Stackelberg strategic game. They are as follows: leaders and followers. The leader moves first and then the follower moves next. Therefore, knowing the leaders move, the follower chooses a strategy to optimize its own objective function (e.g. utility function). This is *lower level*.

By predicting the optimal response of the follower the leader, thus, has to optimize its own objective function (e.g. utility function). This is *upper level* of solution.

As in our power allocation problem CRs choose their requested SINR based on the unit-prices previously announced by BSs, we conclude that the strategies of CRs depend on strategies of BSs that move first. So, the Stackelberg game suits for modeling our problem best[7].

B. SYSTEM MODEL

We consider non-orthogonal multiple access scheme e.g. CDMA applied to a underlay cognitive radio network with a total of M secondary CRs and L channels in a typical cell. There are some assumptions about system model, (1) each channel can be used simultaneously by multiple secondary CRs via some form CDMA access scheme; (2) a single secondary user can use multiple channels at the same time to meet their rate requirements; (3) every active SU radio has an upper limit on power and rate (bits/channel use) at which it can receive; (4) all SUs employ quadrature amplitude modulation (QAM), amplitude shift keying (ASK) and Phase shift keying (PSK) scheme with an adaptable modulation order M; (5) simple path loss model for channel has been assumed; (6) each user has a minimum rate and BER constraint that needs to be maintained and (7) an interference temperature threshold to protect possible primary user transmission on any channel.

The strategy of each BS is to determine the unit-price for each of its CRs and strategy of each CR is to determine the expected SINR level. Each BS tends to earn more utility from the CRs in its cell while spending less transmit power to them. Each CR is interested in gaining higher SINR level (or equivalently data rate) which paying less to the BS of its cell[7]. As a result, the utility functions of BS and CRs can be defined as follows:

$$U_{BS} = \sum_{k=1}^{L} \sum_{i=1}^{N_S(k)} \Pi_i \gamma_i(k) - \sum_{k=1}^{L} \sum_{i=1}^{N_S(k)} p_i(k) \quad (1)$$

Where Π_i is the unit-price determined by *BS* for charging CR_i and γ_i(k) is level of provided received SINR for which is a function of *p* (the vector of powers). The utility function of each CR can be written as:

$$U_{CRi} = \sum_{k=1}^{L} b_i(k) - \Pi_i \sum_{k=1}^{L} \gamma_i(k) \quad (2)$$

where

$b_i(k) = \log(1+\gamma_i(k))$ and

$$\gamma_i(k) = \frac{p_i(k) h_{i,i}(k)}{\sum_{j=1,\ j\neq i}^{N_S(k)} P_j(k) h_{j,i}(k) \rho_{ji}^2 + \sigma^2(k)}, \forall i, k, \quad (3)$$

TABLE I
NOTATIONS

Π_i	Pricing factor of *i*-th user that is broadcasted by BS
$\gamma_i(k)$	SINR per bit for *i*-th user in *k*-th channel
$N_S(k)$	Number of users for *k*-th channel
$\sigma^2(k)$	Noise variance in *k*-th channel
$\rho_{j,i}$	Orthogonality factor between users *j* and *i*
$h_{i,i}(k)$	Power gain from *i*-th transmitter to *i*-th receiver in *k*-th channel
$h_{i,m}(k)$	Power gain from *i*-th transmitter at location m in *k*-th channel
$p_i(k)$	Transmit power per bit of *i*-th user in *k*-th channel
$p_i^{\max}(k)$	Maximu transmit power per bit of *i*-th user in *k*-th channel
$I_{th}(k)$	Interference temperature constraint in *k*-th challel
$b_i(k)$	Rate of *i*-th user in *k*-th channel
$b_i^{\max}(k)$	Maximum rate of *i*-th user in *k*-th channel
$R_{ch}^u(k)$	Maximum rate supported by *k*-th channel
R_i^l	Minimum required rate for *i*-th user
$p_{e,i}(k)$	BER for *i*-th user in *k*-th channel
$p_{e,i}^{th}$	BER threshold at receiver for *i*-th user in any channel

IV. RESOURCE ALLOCATION FRAMEWORK

Power and rate constraints of secondary CRs described in [9] has a centralized optimal solution. But it cumbersome and computationally ineffective. We have converted this to a game which is user based distributed approach. Here the set of power and rate of BS(leader) and CRs(followers) are nonempty, compact and approximately convex. In (1) and (2) both the utility functions are linear which can be considered as concave. Hence we have existence of a Stackelberg NE. It is simple to show that distributed algorithms, which the system can converge to Stackelberg equilibrium. In [10]-[11] the authors showed convergence of CRs to Stackelberg NE using the concept of a potential non-cooperative game.

It is easy to verify that the power allocation played by *CR*s and *BS* is a potential game with potential function in (1) and (2) which they try to maximize,

$$f(x) = \sum_{k=1}^{L} b_i(k) - \Pi_i \sum_{k=1}^{L} \gamma_i(k) \quad (4)$$

$$f(x) = \sum_{k=1}^{L} \sum_{i=1}^{N_S(k)} \Pi_i \gamma_i(k) - \sum_{k=1}^{L} \sum_{i=1}^{N_S(k)} p_i(k) \quad (5)$$

The strategy spaces of CRs and the power game of the *BS* are compact and convex sets and the potential function *f* is continuous. Hence there exists at least one NE for the game for each price Π_i. Also the best response iteration converges to a Nash point[7] and [13].

V.PRACTICAL SOLUTION

A. GAME FORMULATION

The formulation of the proposed Stackelberg game can be decomposed into two levels: lower level of the CRs and upper level for the BS.

1)Lower level: Given the BS's transmit power, the CRs' non-cooperative sub-game can be formulated as strategic non-cooperative game.

In CR networks, the CR transmitters interact with each other. Hence, game theory suits best to analyze the behavior of the system. In game theory, the QoS that cognitive user received is referred to as utility function and cognitive users are the players of the game that finds to maximize their utility. The formal non-cooperative power control game of CR system can be defined as follows.

$$G = [\Omega, \{p_i\}, \{b_i\}, \{U_i(.)\}] \quad (6)$$

where $\Omega = \{1,2,3..............,N\}$ is the players (CRs) index set, $p_i = [0, p_i^{\max}]$ represents the transmission power strategy set of user i, and $p_i^{\max}$ represents the maximum transmission power of user i,. $b_i = [1, b_i^{\max}]$ represents the transmission bit rate strategy set of user i, and $b_i^{\max}$ represents the maximum transmission bit rate of user i. Conventionally, a single parameter formulates the search space. But in this game the search space constitutes two parameters and the whole space is the product of power and rate. For ease in presentation, we define the action for user i, as

$$y_i = [p_i^T \; b_i^T]^T,$$

where, $(\mathrm{p}_{\mathbf{i}} = [p_i(1)\; p_i(2).....p_1(L)]^T$ and.

$$\mathrm{b}_{\mathbf{i}} = [b_i(1)\; b_i(2).....b_1(L)]^T)$$

We consider utility function of user *i* as in (2)

$$u_i(y_i y_{-i}) = \sum_{k=1}^{L} b_i(k) - \Pi_i \sum_{k=1}^{L} \gamma_i(k) \quad (7)$$

where y_{-i}, is the union set of all other users actions and $y_{-i} = [y_i^T \ldots y_{i-1}^T \; y_{i+1}^T \ldots y_M^T]^T$. The non-cooperative game formulation to determine transmit power and rate can be formally stated as

Determine y_i

To Maximize $u_i(y_i, y_{-i})$

Subject to

$$CG1: 0 \le p_i(k) \le p_i^{g\max}(k), \quad \forall\, i,k$$

$$CG2: 1 \le b_i(k) \le b_i^{g\max}(k), \quad \forall\, i,k$$

$$CG3: \sum_{k=1}^{L} b_i(k) \ge R_i^l, \quad \forall\, i,$$

$$CG4: -\gamma_i(k) \le -C_{q\arg}(2^{b_i(k)} - 1),$$

$$\forall\, i,k. \quad (8)$$

From the resource allocation framework [6] the system constraints formulate non-cooperative game. We take a conservative approach to satisfy those constraints in the game formulation.

First assumption the total interference caused by all CRs in a channel is divided equally across all CRs in that channel. This approach results in changing maximum limit on transmit power for each CR. In (13), this is captured in constraint $CG1$. In the game G, maximum limit on power

$$p_i^{g\max}(k) = \mathrm{Min}(\, p_i^{\max}(k),\ Upper\_bound\_p) \quad (9)$$

and the upper bound obtained from bound corresponds to $I_{th}(k)/(h_{i,m}N_s(k))$ for location m. As an example using Table V

$$p_i^{g\max}(k) = \mathrm{Min}(6, I_{th}(k)/(h_{i,m}N_s(k))) \quad (10)$$

Second assumption total supported rate in a channel is also divided across all CRs in that channel. This approach results in changing maximum limit on possible rate for each CR. In (8) this is captured in constraint $CG2$ in the game G, the upper bound on maximum limit on rate

$$b_i^{g\max}(k) = \mathrm{Min}((b_i^{\max}(k)),\ Upper\_bound\_b) \quad (11)$$

and the upper bound obtained from bound corresponds to $R_{ch}^{u}(k)/(N_S(k))$. As an example using Table V

$$b_i^{g\max}(k) = \mathrm{Min}(5, R_{ch}^{u}(k)/(N_S(k))) \quad (12)$$

Here $C_{q\arg}$ calculated from the following BER formula for the modulation scheme using Table V

for QAM

$$P_{e,i}(k) \le \frac{4}{b_i(k)} Q\left(\sqrt{\frac{3b_i(k)\gamma_i(k)}{(2^{b_i(k)} - 1}}\right), \forall i,\, k, odd\ b_i(k); \quad (13)$$

$$P_{e,i}(k) = \frac{4}{b_i(k)}\left(1 - 2^{-\frac{b_i(k)}{2}}\right) Q\left(\sqrt{\frac{3b_i(k)\gamma_i(k)}{(2^{b_i(k)} - 1}}\right),$$

$$\forall\, i,\, k,\, even\ b_i(k);$$

for ASK

$$P_{e,i}(k) = \frac{2}{b_i(k)} Q\left(\sqrt{\frac{6b_i(k)\gamma_i(k)}{(2^{2b_i(k)} - 1}}\right), \forall i,\ k;$$

for PSK

$$P_{e,i}(k) = \frac{2}{b_i(k)} Q\left(\sqrt{\frac{2\pi^2 b_i(k)\gamma_i(k)}{2^{2b_i(k)}}}\right), \forall i,\ k;$$

Here, $Q(x)$ is defined as $\int_x^{\infty} e^{-\varsigma^2/2} d\varsigma$

In summary, based on local information, in order to minimize total power consumption, maximize rate, and maintain QoS(BER), we formulate a game to determine suboptimal distribution of power and rate that a secondary user has to employ across the channels.

2) Upper level: For the BS, if it can be CRs' expected reactions to its action, we are to maximize the utility functions (1) of BS.

B. ANALYSIS OF THE GAMES

Algorithm 1 Algorithm to reach Stackelberg NE for the game *G*

Input: $t_{\max}$, $\in$,*M*, *L*, $p_i^{\max}(k)$, $b_i^{\max}(k)$, $p_{e,i}^{th}$ and $R_{ch}^{u}(k)$;

Output: p_i^t ;

Stopping counter, *t* = 1;

while $(t \le t_{\max} or(\left\| p_i^t - p_i^{t-1} \right\| / \left\| p_i^{t-1} \right\| \le \in), \forall i)$ *do*

% *Execute optimization problem*

for $i = 1, \quad 2,..., M \ do$

for $k = 1, \quad 2,..., L \quad do$

Measure the interference and noise power

$(i.e., \Sigma_{j=1,\ j \ne i}^{N_S(k)} p_j^{t-1}(k) h_{j,i}(k) \rho_{j,i}^{\ 2} + \sigma^2(k))$

across the intended channels;

end for

Solve optimization problem(8) and obtain p_i^t and b_i^t;

end for

for $i = 1,2,..., M \ do$

Transmit p_i^t;

end for

$t = t + 1;$

end while

Comparing distributed approach with centralized scheme we find that the user-based distributed approach is more attractive than centralized scheme in terms of information exchange requirement. As the centralized scheme requires information about all users and channels in the network it incurs a high communication overhead and poor scalability in CRN with large number of CRs. The required amount of information exchange in centralized scheme is *O*(M^2). As a result, whereas, in the developed user-based distributed approach, each CR requires only local information- (i) possible number of users at next time instant $N_s(k), \forall k$ and measurement of interference and noise power.

TABLE II

SPECTRUM USAGE PATTERN ACROSS CHANNELS

Channel, *K*	1	2	3	4	5	6	7	8	9	10	11
User, 1	1	1	1	0	0	1	1	0	0	0	1
User, 2	1	0	1	0	0	1	1	0	0	0	1
User, 3	0	1	1	0	1	1	0	0	1	0	1
User, 4	0	1	1	0	1	1	1	0	1	1	1
User, 5	0	1	1	1	1	1	1	1	1	1	1
User, 6	0	1	1	1	1	0	1	1	1	1	1
User, 7	0	1	0	1	1	0	1	1	0	1	1
User, 8	1	1	1	0	1	0	1	1	0	1	1
User, 9	1	0	1	0	1	0	1	1	0	1	1
User, 10	1	0	1	0	1	1	0	1	0	1	1

TABLE III

CHANNEL QUALITY PARAMETERS

Channel, *K*	1	2	3	4	5	6	7	8	9	10	11
$\sigma^2(k),(\times 10^{-3})$	6	5	4	3	3.5	7	5	5	6	4.5	5.5

VI. NUMERICAL RESULTS

We study the performance of lower level of Stackelberg game which is a non-cooperative game of cognitive radio network. In the simulations, we assume that there are *L* = 11 available channels and a total of *M* = 10 secondary users. Table III provides information on the channel quality for all *L* channels. Table IV lists the minimum rate requirement for each SU [4]. Finally table V contains all other system parameters that are relevant to our resource allocation framework.

TABLE IV

MINIMUM RATE REQUIREMENT OF USERS

User, *i*	1	2	3	4	5	6	7	8	9	10
R_i^l	9	4	5	13	10	8	14	6	11	9

Table V
SYSTEM PARAMETERS

$\Pi_i^{\max} \forall\, i$	0.1
$p_i^{\max}(k)\ \forall\ i,\ k$	5
$b_i^{\max}(k)\ \forall\ i,\ k$	6
$p_{e,i}^{th}\ \ \forall\ i$	10^{-3}
$I_{th}(k)\ \forall\ k$	$200\times\ \sigma^2(k)$
$R_{ch}^{u}(k)\ \forall\ k$	36
$p_{j,i}$	0.03125
$t_{\max}$	10
ε	0.01

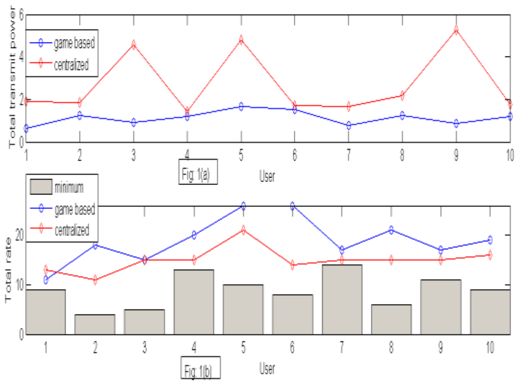


Fig 1. Allocation of total transmit power and total rate of CRs for ASK..

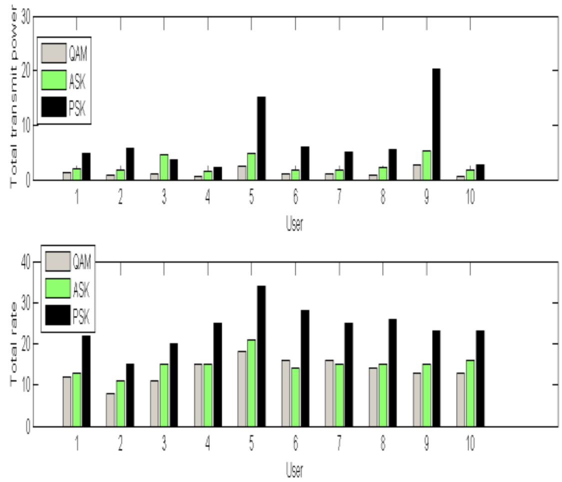


Fig 2. Comparison of total transmit power and total rate for Game Theory of CRs for QAM, ASK. and PSK respectively.

Based on all this information, our objective is to find the optimal transmit power and rate that each of the CRs should employ to guarantee their QoS(BER) through our user-based distributed approach.

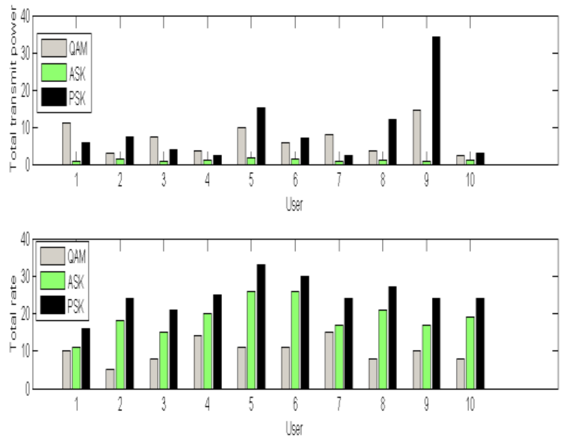


Fig 3. Comparison of total transmit power and total rate for Centralized Scheme of CRs for QAM, ASK. and PSK respectively

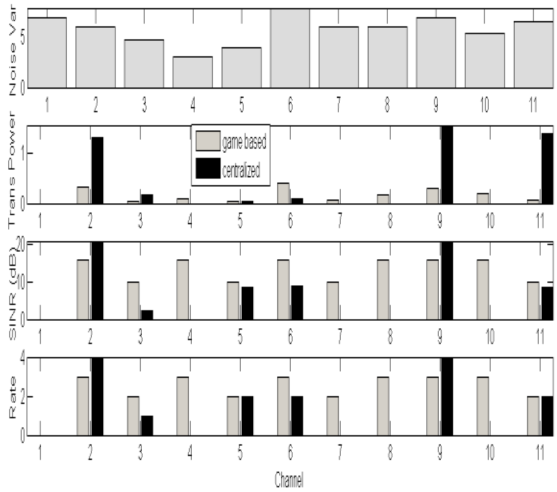


Fig 4. Allocation of transmit power and rate with channel noise, variance and SINR for ASK.

Fig. 1(a) shows the allocation of total transmit power across users from both centralized from paper [9] and distributed (8) schemes, respectively. Fig. 1(b) shows the allocation of total rate across users from both centralized from paper [9] and distributed (8) schemes respectively. We see from Figs. 1(a) and 1(b) that both total allocated power and rate across users in distributed case are comparable to centralized scheme. Because in the game (8) maximum limit on power is set as the minimum of maximum usual limit on power and the upper bound. The upper bound can be obtained from dividing interference temperature threshold by the product of channel power gain at some location and possible number of users at next time instant as explained in (10). For the above approximations distributed solution gives less power than the counterpart. Additionally, our proposed distributed resource allocation scheme is successful in meeting minimum rate requirements for all CRs. The reason is obvious from the proposed user-based optimization problem formulation (8) after checking the feasibility of the optimization problem solution. The feasibility is determined by user minimum rate requirement (constraint ($CG3$). For each user, if the optimization problem is feasible, the distributed scheme is guaranteed to be successful in meeting the rate requirements for all CRs. Fig 1(b) clearly indicates that the proposed game statistics ensure the minimum rate requirement for all users.

Fig 2 presents the of total transmit power across users of distributed scheme for QAM, ASK and PSK modulation. From the constant C_{qarg} calculated in [3] for QAM, ASK and PSK, we can predict that the QAM has the lowest power allocation and the PSK has the highest power allocation. Consequently the rate of three modulation schemes have the similar properties. Similar representation for centralized scheme is shown in Fig-3.

Fig.4 presents the transmit power and rate allocation across channels for user 5 from the proposed distributed scheme(8) along with centralized scheme in [9]. The channel noise variance and resulting SINR are also shown for reference. User 4 operates on channels 2, 3,4, 5, 6, 7,8, 9, 10 and 11. A game theory does not always give optimal result. We can observe the effect in the channel 6. Here game theory requires more transmit power.

Fig. 5(a) and 5(b) show the resulting total interference power and allocation of total rate, respectively, across channels from distributed scheme along with upper limits. We see from Figs 5(a) and 5(b) that both resulting total interference power and total rate across channels do not violate the corresponding upper limits. That is, the conservative approach based on constraints $CG3$ and $CG4$ in the proposed distributed case is successful in satisfying the system constraints in [9].

Upper limits in channels are used for fairness. The main purpose is to satisfy the requirement of individual users which is shown in fig: 1 to be satisfied.

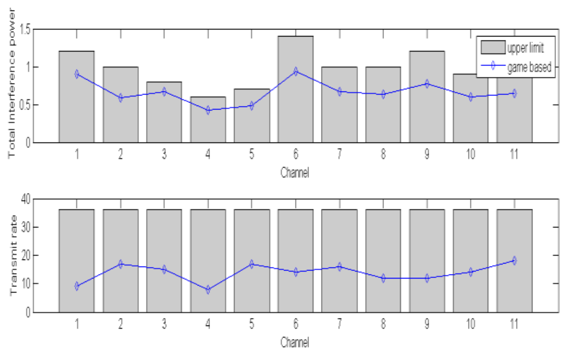


Fig. 5. Total interference power and total rate across channels

VII. CONCLUSION

Nowadays Game theory based distributed approach is used for its easier implementability. In this paper, we consider the problem of power allocation to maximize utilities of users (BS and CRs) in a cognitive radio network that is suitable for Stackelberg Game. Using game theory, we propose a distributed power control scheme to protect primary users from interference while providing maximum possible utilization of base stations and cognitive radios. In addition, the received SINR level of CRs determined by the level requested by each individual CR. The proposed utility functions have led to a distributed architecture which is more realistic model. Numerical results are presented to prove the effectiveness of the proposed distributed algorithm.

REFERENCES

[1] G. Scutari, D.P. Palomar, F. Facchinei, and J.S. Pang, "Convex Optimization, Game Theory, and Variational Inequality Theory," IEEE Signal Processing Magazine, vol. 27, no. 3, pp. 35-49, May 2010.

[2] T. Zhang, W. Chen, Z. Z. Cao "Hierarchic Power Allocation for Spectrum Sharing in OFDM-Based Cognitive Radio Networks" Vehicular Technology, IEEE Transaction vol. 63,no.8, Oct. 2014.

[3] L.Akter, B Natarajan, "QoS Constrained Resource allocation to Secondary Users in Cognitive Radio Networks."Computer Communication on Cognitive Radio and Dynamic Spectrum Sharing Systems." June 2009.

[4] X. Huang, B. Beferull-Lozano, "Non-Cooperative Power Allocation Game with Imperfect Sensing Information for Cognitive Radio", IEEE International Conference on Communications on Cognitive Radio and Networks Symposium, 2012.

[5] O. E. Ferkouss, W. Ajib," Game theory based resource allocation for cognitive radio networks",Globecom on on Cognitive Radio and Networks Symposium, 2012.

[6] X. Huang, B. Beferull-Lozano, C. Botella," Quasi-Nash Equilibria for Non-convex Distributed Power Allocation Games in Cognitive Radios", IEEE Transactions on Wireless Communications, , vol. 12, no. 7, pp.2326-2339, July 2013.

[7] N. N. Omidvar, B. H. Khalaj. "A Game Theoretic Approach for Power Allocation in the Downlink of Cognitive Radio Networks," in IEEE International Conference on Computer Aided Modeling and Design of Communication Links and Networks (CDMAD)." 2011.

[8] E. Hossain, L. Le, N. Devroye and M. Vu. Cognitive Radio: From Theory to Practical Network Engineering., New Directions in Wireless Communications Research. pp 251-289, August 2009.

[9] L. Akter, B. Natarajan. "Game Theory based Distributed Approach for Power and Rate Allocation to Secondary Users in Cognitive Radio Networks", International Conference on Informatics, Electronics & Vision (ICIEV)." May 2013.

[10] X Kang, Y-C Liang, H K Garg, "Distributed Power control for Spectrum-Sharing Femtocell Networks using Stackelberg game", IEEE International Conference on Communications, 2011.

[11] V.U. Sankar and V. Sharma, "Pricing and Power allocation in Femtocells using Stackelberg Game Theory", IEEE International Conference on limit Signal Processing and Communications (*SPCOM*), 2014.

[12] S. M. A.Karim, S. Mohiyan and L. Akter" Allocation of Power and Modulation Order (Rate) to Secondary Users in Competitive Cognitive Radio Networks", International Conference on Informatics, Electronics & Vision (ICIEV)." May 2014.

[13] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

Comparative Analysis of Protein Alignment Algorithms in Parallel environment using CUDA

BLAST versus Smith-Waterman

Shadman Fahim
Department of Computer Science and Engineering
BRAC University
Dhaka, Bangladesh
shadmanbracu09@gmail.com

Shehabul Hossain
Department of Computer Science and Engineering
BRAC University
Dhaka, Bangladesh
rudrozzal@gmail.com

Gulshan Jubaed Prince
Department of Computer Science and Engineering
BRAC University
Dhaka, Bangladesh
jubaedprince@hotmail.com

Jia Uddin
Department of Computer Science and Engineering
BRAC University
Dhaka, Bangladesh
jia.uddin@bracu.ac.bd

***Abstract*—In bioinformatics to identify evolutionary relationships two sequences are matched to find similarities. Smith Waterman, a dynamic algorithm, is a common choice to carry out this alignment process. However, with the exponential growth of protein databases this algorithm's time complexity increases. The demand of bioinformatics for their tasks to speed up is very high. Even a slight seed up in computation would be very helpful in the field of bioinformatics. Thus, for a lot of the scientists this algorithm might not be the first choice. In today's world the most popular and used bioinformatics tool is the BLAST (Basic Local Alignment Tool). BLAST, similar to Smith Waterman algorithm, is an alignment algorithm for scanning proteins from protein databases. This paper analyzes both the algorithms in a parallel environment with the help of NVIDIA GPU. For our experiments we utilized a GeForce GTX 660 NVIDIA GPU and env_nr dataset. Experimental results show that parallel implementation of BLAST algorithm mostly range 2-5 times faster than parallel Smith-Waterman.**

Keywords-CUDA, GPU, BLAST, HSP, HSA, NCBI.

I. Introduction

Basic Local Alignment Search Tool (BLAST) is the most popular alignment algorithm in the world of science today. The algorithm uses dynamic programing which utilizes well defined mutation scores. This method is more than an order of magnitude faster than the existing heuristic algorithms [1]. It has been cited over 25,000 [2] and over 21,000 [3]. For such popularity US National Center for Biotechnology Information (NCBI) plays the most important role. NCBI provided a platform over the internet for everyone's reach. Now any general person can go on the website of NCBI and get results for their queries. It is noted that hundreds and thousands of queries are being processed every now and then using the platform provided by NCBI. This increase BLAST's usage by 2 to 3 times [4].

On the contrary, Smith Waterman is also a dynamic algorithm but isn't used as much as BLAST. This algorithm generates more accurate results than what BLAST produces. However, its accuracy is maintained at the expense of computation time and computer power [5]. Computation speed is the burning topic in today's world. How fast a task can be processed is the main challenge. One of such other challenges is searching through long detailed databases. With the exponential growth of protein databases demand to accelerate searching through such huge databases is very high. NCBI is having tremendous breakthroughs in this particular field. However, NCBI uses sequential search for the queries. With the availability of Graphics Processing Units (GPUs) it can be assumed that using its parallel techniques BLAST algorithm can have a faster processing time. An implementation of BLASTP algorithm is handled by GPU using Compute Unified Device Architecture (CUDA), CUDA-BLASTP. It is claimed that in CUDA architecture they have managed to achieve speedups of 10 times compared to sequential NCBI BLAST 2.2.22 on a GeForce GTX 295. It is also 3-4 times faster than multithreaded NCBI BLAST on an Intel Quad-Core processor [4]. Similarly, mpiBLAST is an open-source sequence tool that parallelizes the NCBI BLAST toolkit. It uses the database segmentation approach and the master-worker style. It achieves significant speedups in small or moderate number of processes [6]. On the other hand researchers started to implement Smith-Waterman in GPUs as well. CUDASW++ 2.0 has managed to achieve an average performance of 9.509 GCUPS on single-GPU version and an average performance of 14.484 GCUPS on dual-GPU version [5].

Hence there are scopes to speed-up the process and meet the demand of accelerating it. This paper demonstrates both the alignment algorithms. Also, it illustrates how this task is handled by a GPU using CUDA. CUDA by NVIDIA, a parallel computing architecture uses parallel compute engine in NVIDIA GPUs to solve many computationally intensive problems in a more efficient way than on Central Processing Unit (CPU) [7]. Using its parallel techniques we demonstrate how computation can speed-up. The database used in this research is taken from the NCBI website ftp://ftp.ncbi.nlm.nih.gov/blast/db/FASTA/env_nr_gz.

The remainder of the paper is organized as follows - Section II displays the architecture of a NVIDIA GPU, features of CUDA, Smith-Waterman algorithm and BLAST algorithm. Section III provides a detailed explanation of the implementation of the algorithms. The results of the experiments carried out and their analysis are included in Section IV. Finally Section V concludes the paper.

II. BACKGROUND STUDY

A. *GPU Architecture and CUDA*

GPU is considered to be efficient and have a better performance. Comparing to a CPU, a GPU provides a better performance because it offers a higher peak GFLOPS (Giga floating-point operations per second) [8]. The GPU that we used for the experimentations is GeForce GTX 660. Generally a GPU device has several multiprocessors with several processors inside each of them. Figure 1 enlightens it. There are mainly two types of memory in GPU. One is on-chip memory and the other is off-chip memory. The on-chip memory has low access latency but a relatively small size. On the other hand the off-chip memory has larger size and also higher access latency [9]. Moreover, these microprocessors contain the shared memory and caches, along with registers.

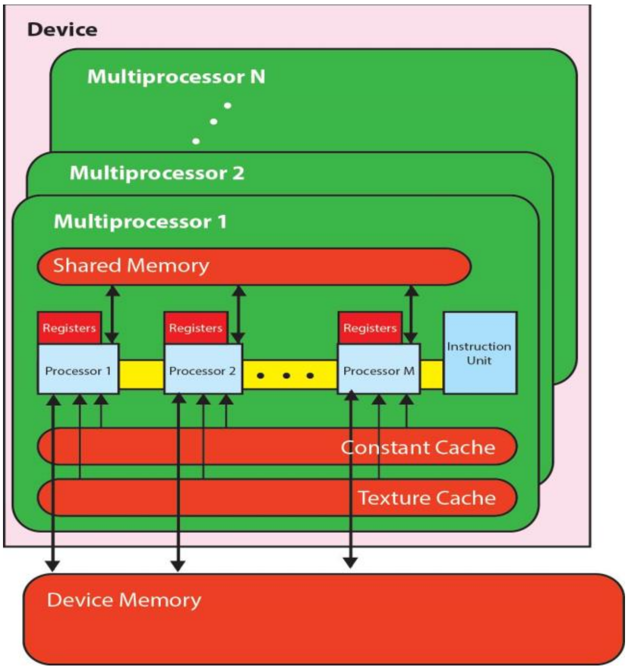


Figure 1. General architecture of a GPU.

CUDA introduced by NVIDIA is a general purpose parallel computing platform and programming model [10]. The CUDA functions are called kernels. Unlike C functions that run only once this kernel runs N times in parallel by N different CUDA threads. Each thread that executes the kernel is given a unique thread ID. The threads are organized in a hierarchy consisting of blocks and grids. When calling a kernel function the size of the blocks and the number of threads per block are specified. An example of the function to call kernels is presented as:

kernel<<<numBlocks, numThreads >>>(parameter's list).

B. *BLAST Algorithm*

Before BLAST, FASTA was developed by David J. Lipman and William R. Pearson in 1985 [11]. Besides fast algorithms like BLAST and FASTA, Smith-Waterman algorithm was used to search protein databases which guarantee the optimal alignments of the query and database sequences unlike BLAST and FASTA. However, the heuristic approach of BLAST algorithm is overall a lot faster. So, due to such highly populated protein databases Smith-Waterman search is both time consuming and computer power intensive.

So as mentioned earlier BLAST is the most popular heuristic search algorithm for protein scanning. Unlike Smith-Waterman algorithm where the entire sequence is compared BLAST locate high scoring short matches between the query sequence and the subject sequence [9]. Due to this the accuracy of BLAST decreases to some extent but then the processing speed increases exceptionally than Smith-Waterman. The Blast algorithm mainly has four stages [12]. The first stage is the hit detection where the query sequence is matched with the subject sequence in order to find matches. The query sequence is broken down to user defined size word lengths (W). Then the words are compared with subject sequences to detect similarities. The hits are then scored using the blosum62 scoring matrix. A sample snap shot of blossom62 is presented in Figure 2.

	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V
A	4																			
R	-1	5																		
N	-2	0	6																	
D	-2	-2	1	6																
C	0	-3	-3	-3	9															
Q	-1	1	0	0	-3	5														
E	-1	0	0	2	-4	2	5													
G	0	-2	0	-1	-3	-2	-2	6												
H	-2	0	1	-1	-3	0	0	-2	8											
I	-1	-3	-3	-3	-1	-3	-3	-4	-3	4										
L	-1	-2	-4	-4	-1	-2	-3	-4	-3	2	4									
K	-1	2	-1	-1	-3	1	1	-2	-1	-3	2	5								
M	-1	-1	-3	-3	-1	0	-2	-3	-2	1	2	-1	5							
F	-2	-3	-3	-3	-2	-3	-3	-3	-1	0	0	-3	0	6						
P	-1	-2	-1	-1	-3	-1	-1	-2	-2	-3	-3	-1	-2	-4	7					
S	1	-1	0	0	-1	0	0	0	-1	-2	-2	0	-1	-2	-1	4				
T	0	-1	-1	-1	-1	-1	-1	-2	-2	-1	-1	-1	-1	-2	-1	1	5			
W	-3	-3	-4	-4	-2	-2	-3	-2	-2	-3	-2	-3	-1	1	-4	-3	-2	11		
Y	-2	-2	-3	-3	-2	-1	-2	-3	2	-1	-1	-2	-1	3	-3	-2	-2	2	7	
V	0	-3	-3	-3	-1	-2	-2	-3	-3	3	1	-2	1	-1	-2	-2	0	-3	-1	4

Figure 2. The scoring matrix, Blossum 62.

The hits are not necessary to be exact matches, similar matches is also accepted as long as the score of that hit is greater than a certain threshold (T). These are then saved in an efficient data structure such as a lookup table.

In stage 2, the remaining hits are now sent to this stage for further processing. The hits are extended in both directions and as long as the accumulated score is increasing the extension is carried on. As soon as the score starts to decrease we stop. This is called un-gapped extension. The result is HSPs (highest scoring pairs). A sample of un-gapped extension is presented in Figure 3.

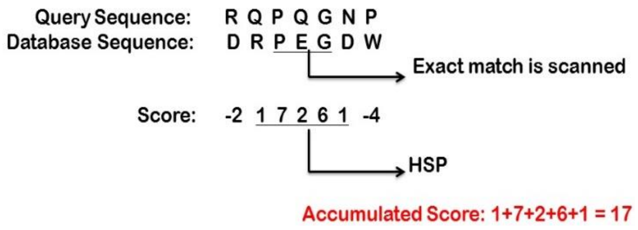


Figure 3. Un-gapped extension.

Consequently in stage 3, the HSPs sent to this stage are then extended further. However, unlike the last stage here gaps are allowed. With each gap there is a penalty. This is called the gapped extension. Finally in stage 4, scores all the alignments again from the previous stage. Once done scoring it produces the top scores. This is called the gapped alignment with trace-back. The entire BLASTP algorithm is enlightened in Figure 4.

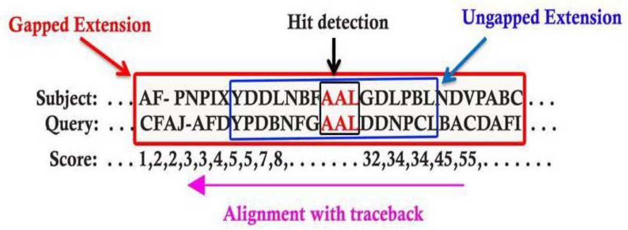


Figure 4. An overview of BLAST algorithm.

In Figure 4, the query sequence and the subject sequence is compared. We can see there is an exact match in the sequence AAL. This is the stage one where the hit is detected. In the next stage the un-gapped extension is performed on pairs of high-scoring segment pairs (HSPs). Here, the initial match in extended in both directions until there the overall accumulated score starts to decrease. In stage 3, the un-gapped extension is extended using a gapped alignment [13]. To determine the level of the alignment a scoring matrix and a threshold value is used. Finally, in the last stage a trace-back algorithm is used to produce and score the alignments.

The time complexity of this particular algorithm is [14]:

$$O(M) + O(MN) + O(1) = O(MN).$$

Where, M is the number of look ups in the hash table to finds seeds and N is the length of the query sequence. Also, because calculating the statistical significance of HSP is a constant time operation, these have a complexity of O(1).

C. *Smith Waterman Algorithm*

In 1981 before BLAST or FASTA were written Smith and Waterman suggested Smith-Waterman algorithm [15]. Later in 1982 Gotoh improved the algorithm [16]. This is a local alignment algorithm. Thus it matches the highest similarities between two proteins instead of the aligning the entire two proteins. Assuming two query sequences S_1 and S_2 having lengths m and n. The two sequences are arranged in a matrix form with m+1 rows and n+1 columns. Initially the first row and column are set to 0. Then the similarity matrix is computed for $1 \le i \le m$, $1 \le j \le n$ using the formula as shown below. At last the trace back is performed to calculate the final overall score.

There are mainly three steps to run this algorithm, they are:

1. Initialization.

M (0, j) = 0;
M (i, 0) = 0; where M is the similarity score matrix.

2. Filling the matrix, M.

$$M(i,j) = max\begin{Bmatrix} 0, \\ M(i-1,j-1)+a, \\ M(i-1,j)-d, \\ M(i,j-1)-d \end{Bmatrix}, 1 < i \le m, 1 < j \le n$$

Where,
a = match/mismatch value
d = gap penalty
m = length of a sequence
n = length of another sequence

3. Trace back the sequences for a suitable alignment.

$F = \max\{M(i,j)\}$;
traceback(F);

The time complexity of this algorithm is [14]:

$$O(M+N) + O(MN) + O(MN) = O(MN).$$

Here M and N are the length of the sequences. O(M+N) is for initialization. For filling the matrix the time complexity is O(MN) whereas for trace back it is O(MN).

III. EXPERIMENTAL SETUP

For our experiments CUDA Toolkit 7.5 is used and NVIDIA GeForce GTX 660. All the experiments are conducted in a personal computer (PC) with the configuration Intel(R) Core i7-4470 CPU @3.4 GHz, 16GB RAM, running Ubuntu 14.04.

A. *BLAST CUDA*

Figure 5 demonstrates a detailed implementation of BLASTP algorithm. It brings light to what part of the code is sent to GPU for execution. Stages 1 and 2 of BLAST algorithm are processed in GPU namely hit detection stage and un-gapped extension. First and foremost the CPU takes the query sequences. Then it sorts the database according to the number of subject sequences it contains. This helps in balancing the load among the threads. So, no threads in the same wrap (cluster of threads that can execute in parallel) work on subject sequences with large length difference. Later, the database is sent to kernel for calculating the HSP pairs. Once done, High Scoring Alignments (HSAs) are computed out in the CPU using gapped extension. At last final calculations are made and the results identical to NCBI-BLAST are displayed.

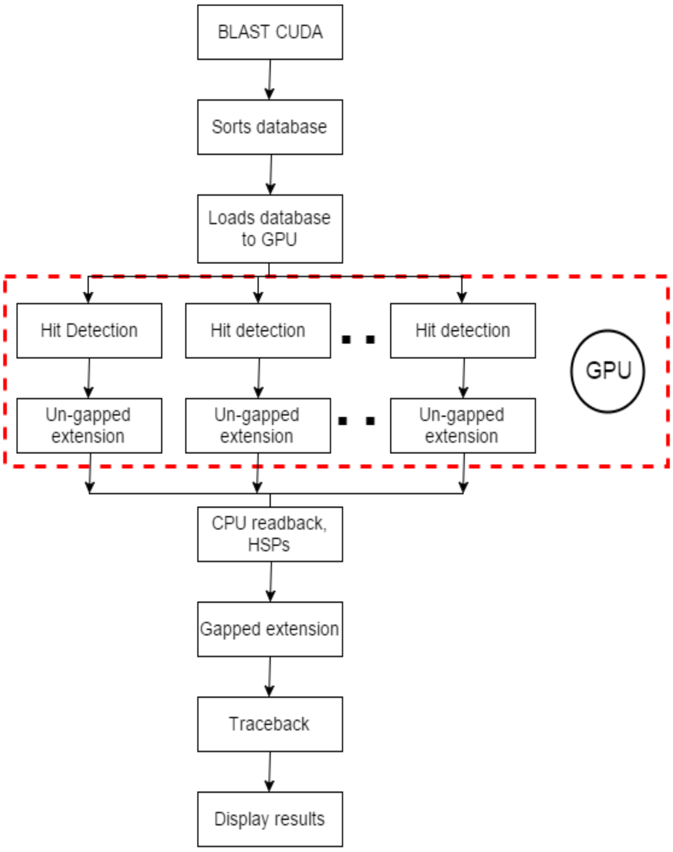


Figure 5. The experimental setup of BLAST algorithm.

- *Input*

The three main inputs of BLAST are the query sequence, database where the subject sequences are stored, and threshold value at which an alignment must score to avoid being cut off.

- *Kernel Call*

Stage 1 and stage 2 are mainly performed in this kernel. At first the query sequence inputted is broken down to several words of length 3. The protein database is stored in the global memory of the GPU. Kernel is then called and the 3 letter words are sent to corresponding threads with a batch of the database. There each word is matched with the subject sequence. Whenever a match is confirmed un-gapped extension is carried out. This generates HSPs. Finally, the HSPs are read back to the host (CPU) for further processing. An overview of the first kernel is enclosed in red dotted box in figure 5.

- *CPU Readback*

The HSPs are read backed to the CPU. Here the later part which is the gapped extension is processed. The results of gapped extension are HSAs. HSAs are filtered if they fail to overcome the threshold value. Finally, after trace backing the final results are outputted on the display.

B. *Smith-Waterman CUDA*

This algorithm is very time consuming as matrices are generated against every subject sequence from the database. Thus running this algorithm in GPUs is preferable as GPUs are designed to compute matrices. The implementation of smith-waterman algorithm is illustrated in Figure 6.

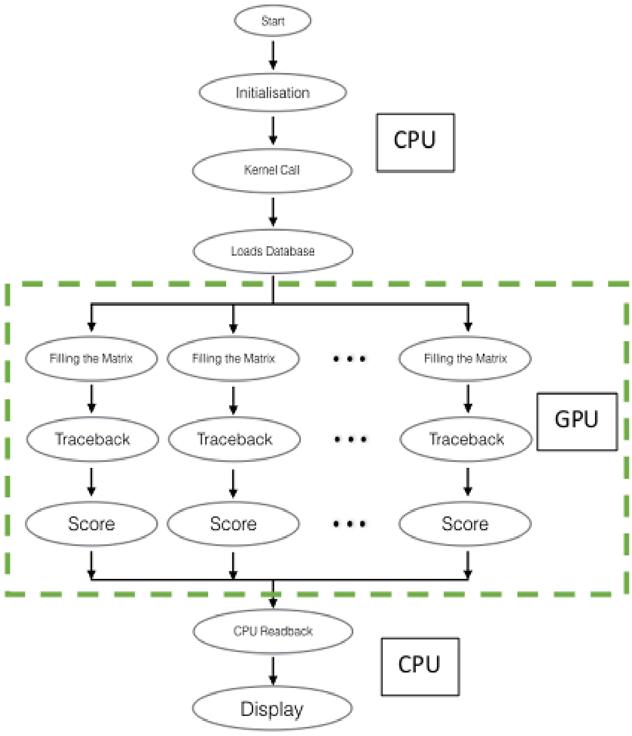


Figure 6. The experimental setup of Smith-Waterman Algorithm.

The operations in doted blocks are carried out by GPU and the rest of the blocks are performed by CPU. Similar to the BLAST CUDA implementation the database is sorted so no threads on the same cluster work on subject sequences with large length difference. Each thread in the GPU is assigned to fill in the similarity matrix against one sequence from the database. Once done matching the similarity between the sequences the matrix is saved in the local memory. Then the third step of the algorithm is performed which is trace back. First the thread figures out the maximum value in the matrix and starts tracing back till it reaches zero. At last along the line of the trace back the alignment found is scored. Then the alignments with scores are read back to CPU where it organizes the results and displays it.

IV. Experimental Results Analysis

The env_nr database we used is of 1.5 GB. Our input query sequences are of yeast and that too retrieved from the NCBI website. The database has a total of 6,891,928 sequences; 1,364,236,057 letters. We varied the query length sequence from 26 to 1002.

At first we carried out a test to figure out the optimal block and thread size to carry out our experiments. Firstly, we kept block size constant and varied the thread size. Then we changed the block size and completed the task again.

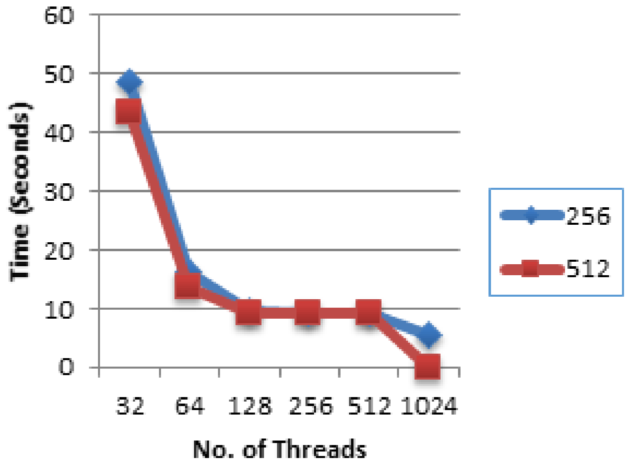


Figure 7. Comparison of execution time varying block and thread sizes.

To find the optimal thread and block sizes we used the Ubp6p protein sequence which is of length 499. The graph generated is shown in Figure 7. For the last value which is marked as X, the GPU we used runs out of global memory to finish the task. GeForce GTX 660 has a memory space of 2 GB. Thus a GPU with a higher global memory will give the result. Similarly when we tried the same job with a larger sequence of length 1002 any thread size greater than 256 shows the unavailability of memory space. Finally, we conclude to complete the experiments with keeping the block size constant at 256 while changing the thread size twice 128 and 256.

According to our results BLAST performs much faster than smith-waterman algorithm. When using 256 block size and 128 numbers of thread on each block BLAST is around 2-5 times faster than smith-waterman shown in figure 8. On the other hand in figure 9 when the block and thread sizes are changed to 256 and 256 respectively BLAST perform with a speed of 1.5-4.5 times faster.

TABLE 1. EXECUTION TIME OF UBP6P PROTEIN IN BLASTP. THREAD SIZE IS VARIED WITH BLOCK SIZE.

Thread size	Time for block size 256	Time for block size 512
32	48.709	43.422
64	16.131	13.495
128	9.538	9.141
256	9.235	9.154
512	9.201	9.133
1024	5.505	X

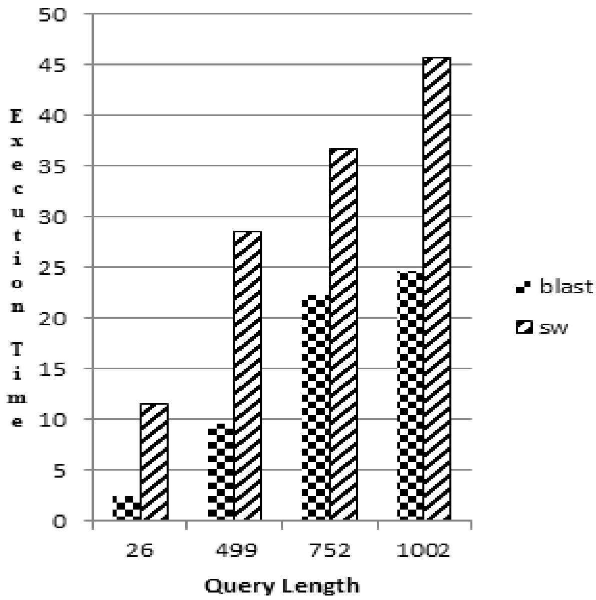


Figure 8. Comparison of runtimes of both the algorithm using Table 2.

TABLE 2. RUNTIME IN SECONDS FOR BOTH THE ALGORITHMS USING 256 GPU BLOCKS AND 128 THREADS.

Query Sequence Length	GPU blocks	GPU threads	BLAST CUDA	SW CUDA
26 (SCY_4187)	256	128	2.432	11.615
499 (Ubp6p)	256	128	9.699	28.482
752 (Gcn20p)	256	128	22.370	36.799
1002 (SAP155)	256	128	24.585	45.749

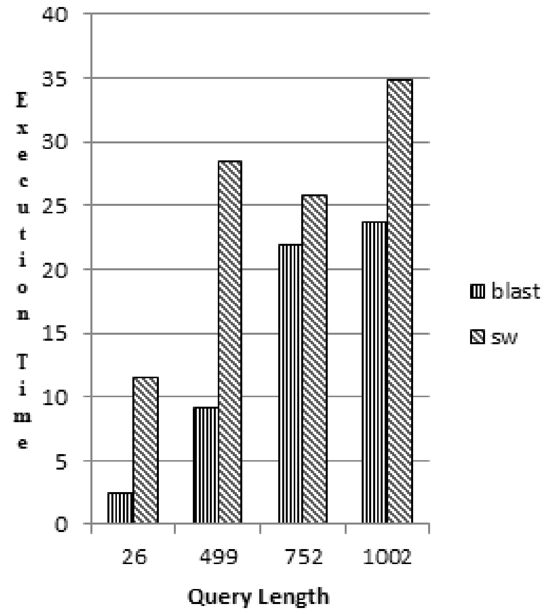


Figure 9. Comparison of runtimes of both the algorithm using Table 3.

TABLE 3. RUNTIME IN SECONDS FOR BOTH THE ALGORITHMS USING 256 GPU BLOCKS AND 256 THREADS.

Query Sequence Length	GPU blocks	GPU threads	BLAST CUDA	SW CUDA
26 (SCY_4187)	256	256	2.446	11.531
499 (Ubp6p)	256	256	9.185	28.420
752 (Gcn20p)	256	256	21.963	25.764
1002 (SAP155)	256	256	23.773	34.821

TABLE 4. RUNTIME OF BLASTP IN SECONDS FOR CPU AND GPU.

Query Sequence Length	BLAST 128	BLAST 256	BLAST CPU
26 (SCY_4187)	2.432	2.446	4.189
499 (Ubp6p)	9.699	9.185	27.474
752 (Gcn20p)	22.370	21.963	45.494
1002 (SAP155)	24.585	23.773	54.652

Since BLAST is the most used algorithm we did more experimenting with it. We ran the entire algorithm in CPU and compared the results with that obtained using GPU. The results are portrayed in figure 10. BLAST 128 means block size 256 and thread size 128. While BLAST 256 means block size 256 and thread size 256.

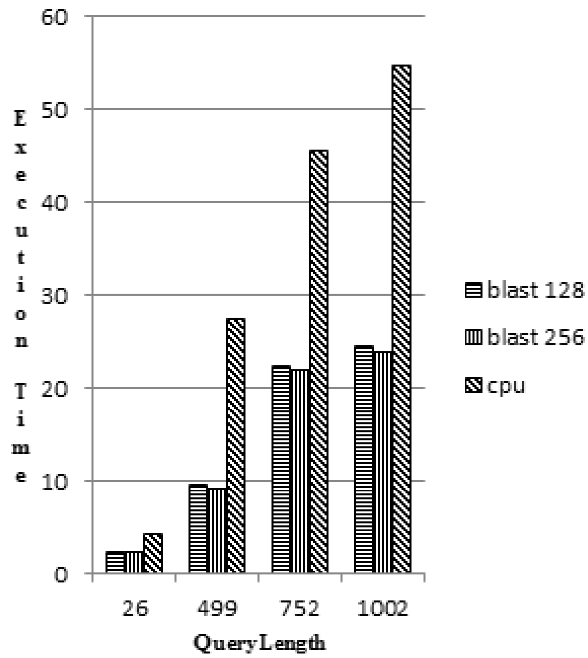


Figure 10. Comparison of runtimes achieved by CPU and GPU using Table 4.

V. CONCLUSION

This paper examined how much BLASTP algorithm and Smith-Waterman varies from each other in computation time using the parallel techniques of CUDA. We have collected the database of protein from NCBI and CUDA Tool kit 7.5 is used. It is 2-5 times faster than the BLAST. Smith-Waterman being a very exhaustive algorithm while performing in CPU is being handled pretty well in GPU. It still gives good results compared to BLAST CUDA. However, falls short in execution time. On the bright side, smith-waterman gives more accurate results than BLAST. Thus, while choosing which algorithm for their alignment task one has to decide based on accuracy or execution time. One of the limitations of our research is that we failed to record the time of Smith Waterman in CPU. The CPU we used were not powerful enough to process matrices of all the sequences in the database. Nevertheless, we hope our results will motivate others to work on GPUs because in today's world being fast is important.

REFERENCES

[1] S. Altschul, "Basic Local Alignment Search Tool," J. of Molecular Biology, vol. 215, no. 3, pp. 403-410, 1990.

[2] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, and D.J. Lipman, "Basic Local Alignment Search Tool," J. Molecular Biology, vol. 215, pp. 403-410, 1990.

[3] S.F. Altschul, T.L. Madden, A.A. Scha¨ffer, J. Zhang, Z. Zhang, W. Miller, and D.J. Lipman, "Gapped BLAST and PSI-BLAST: A New Generation of Protein Database Search Programs," Nucleic Acids Research, vol. 25, no. 17, pp. 3389-3402, 1997.

[4] L. Weiguo, B. Schmidt, and W. Muller-Wittig, "CUDA-BLASTP: Accelerating BLASTP on CUDA-Enabled Graphics Hardware," IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 8, no. 6, pp. 1678-1684, Nov.-Dec. 2011.

[5] L. Yongchao, L.M. Douglas, and S. Bertil, “CUDASW++ optimizing Smith-Waterman sequence database searches for CUDA-enabled graphics processing units,” BMC Reseach Notes, vol. 2, no. 73, 2009.

[6] Lin, Heshan et al. “Coordinating Computation And I/O In Massively Parallel Sequence Search,” in IEEE Transactions on Parallel Distributed Systems, vol. 22, no. 4, pp. 529-543, 2011.

[7] Cheng, John, Max Grossman, and Ty KcKercher. Professional CUDA C Programming. Indianapolis: Wrox, 2014. Print.

[8] NVIDIA GTX 680 Whitepaper. Retrived June 5,2012 from http://www.nvidia.es/content/PDF/product-specifications/GeForce_GTX_60_Whitepaper_FINAL.pdf .

[9] Z. Jing, W. Hao, L. Heshan, and F. Wu-chun, “cuBLASTP: Fine-Grained Parallelization of Protein Sequence Search on a GPU,” in Parallel and Distributed Processing Symposium, 2014 IEEE 28th International , vol., no., pp.251-260, 2014.

[10] NVIDIA Corporation. NVIDIA CUDA C Programming Guide, 2015. Version 7.0.

[11] DJ Lipman, and WR Pearson, “Rapid and sensitive protein similarity searche,” vol. 227, pp. 1435-41.

[12] S. Xiao, H. Lin, and W. Feng, “Accelerating Protein Sequence Search in a Heterogeneous Computing System,” 2011 IEEE International Parallel & Distributed Processing Symposium, IPDPS, pp. 1212 - 1222.

[13] A. David, “An Analysis of BLASTP Implementation on NVIDIA GPUs,” 2012.

[14] A. Chan, “An analysis of pairwise sequence alignment algorithm complexities: Needleman-wunsch, smith-waterman, fasta, blast and gapped blast,” 2007.

[15] T.F. Smith, and M.S. Waterman, “Identification of common molecular subsequences,” J. Mol. Biol, vol.147, pp.195-197, 1981.

[16] O. Gotoh, “An improved algorithm for matching biological sequences,” J. Mol. Biol, vol. 162, pp. 705-708, 1982.

CHAAR:A Location Based Product Offer Advertisement App

S.M. Mohi-Us Sunnat, Amitabha Chakrabarty
Department of Computer Science and Engineering
BRAC University
45 Mohakhali, Dhaka-1212, Bangladesh
Email: suncha629@gmail.com, amitabha@bracu.ac.bd

***Abstract*—Shopping is a major part of this modern era for both Men and Women. CHAAR[1] App dynamically help the users to find out their desire and favorite products. Besides it will give the notification about the shops which will have sale when users are around that shops or market. Moreover, this App will show the route between the users location and the market by using Google Map. The Seller can promotion their products which is in sale or new in the market so that the buyers can easily track their favorite products. In addition, it will help the users to save their time and energy for shopping.**

Keywords: Location Based service, GPS, Google Map, Android, Mobile App.

I. Introduction

The booming growth of the mobile devices[2] taught us the importance of staying connected anywhere any place. Now anyone can be reached around the globe if that individual is carrying a mobile device having voice and/or data connection enabled. The advancement of technology made all this possible. Today's mobile phones are equipped will functionalities that are much beyond voice and SMS. Recent mobile phones are much more powerful than some of the personal computers (PCs) may be 10/12 years back and enabled with the capabilities running much more calculation intensive applications (Apps). Along with the functionalities of data services (such as email, browsing,etc) there are another service that modern mobile phones can provide such as sense of device location. With built in receiver, mobile phones access global positioning system (GPS[3]) satellites to detect it's location accurately (around 10/ 50 meters error). This feature opens up a new domain called location based services (LBSs) [1] where a user can have information based on the location of the mobile phone. When initially introduced, LBS where limited to product advertisement, such as promotional offers in a market place or in airport lounge etc. New and emerging LSB enabled application applications are in huge demand from the user domain [2]. In emergency situation when the current location is very important to pinpoint the affected area, location based information or service comes to rescue. System such as alerting the fire brigade in an event of a fire, location based service allows the fire fighters to locate the victims. Similarly in the need of emergency blood, a blood bank can locate potential individuals of particular blood groups using location based service [4].

In this paper, we are proposing use of LBS in the area of mobile advertisement[5]. We proposed a framework for an android App that will enable user to sense the markets nearby him/her. And the app will also notify user of any promotional offer from the shops in the nearby markets. Promotional offers will be notified based on user interest. Our prototype currently supports Dhanmondi area and near by shopping malls. But can be mapped to the entire Dhaka city or for all over Bangladesh by adding new store locations in the location database.

II. General Idea

This development is based on two very popular things. One is Shopping and another is Geo-Location Based Sale Advertising [8]. We always like to have something extra. If we can get that for shopping that will be always welcomed. There are a lots of local and international brand for cloths which are very popular in our country. During discount period, their selling static increases but most of the buyer are not get the in time information of the discount so that many people miss the offer. Though some people informed by friends, TV ads, social media or local banners, posters etc. But now the customer acceptance for location-based ads is possible due to advancements in mobile technology as well as because companies will be understood that location is only one of targeting dimensions. If the ad will be supposed be effective, it will have to be more relevant, more context oriented, more tailored. We dynamically show a particular market or shop where a retailer promotes an on-sale product.

III. Related Work

There are some similar projects like Location Based Intelligent Advertisement [3]. This is a project where Open Street is used and this project notifies the user about the nearest Markets, Shops, Chain Stores etc.But in our project we are using Google Map and we also do almost similar thing but we are showing the advertisement of on-sale shopping products of the nearest markets or shops. It wont notify the users of a

[1]CHAAR is Bangla for SALE in English
[2]http://data.worldbank.org/indicator/IT.CEL.SETS.P2
[3]http://www.gps.gov/
[4]https://en.wikipedia.org/wiki/Location-based_service
[5]http://advertise.bingads.microsoft.com/en-us/mobile-advertising

market where none on-sale is running. There are other apps where LBS is used for product advertisement [4], [5] This ads targeting is looking at all factors that might influence each consumers purchasing decisions, such as nearby venues, events, the weather, and neighborhood demographics such as the ages and gender, traffic conditions etc.

A. Our Contribution

In this work an android application[6] is designed keeping in view both the consumer as well as the retailer. For advertising any promotional offer, retailers have to use retailer interface for put product(s) advertisement server database. The App would fetch the updated record from the database and displays that to the consumer side. Each of the two parts have been discuss in detail in the following sections.

1) Why Android: In Bangladesh, Android, Apple and Windows Phone is the main Operating System for mobile phones. But Android is the main OS [6] in our country. More than 70% mobile users use Android. Fig 1 show the pie chart of the Android user statistics. So this platform is the best way to reach the users. Besides, this platforms devices are so available and start from very cheap budget. As user experiences, android is very user friendly. So in every sector android is better than other platform in Bangladesh and thats why I choose this platform.

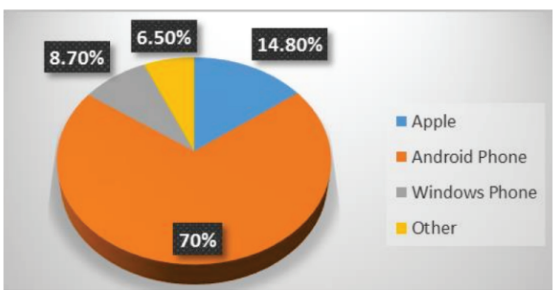


Fig. 1: Android phone use in Dhaka City[6]

B. Database Description

This section highlights the database used in the development of this App. This database has two parts, one is the retailer part and the other one is the consumer side. These two parts are separated using tables designed to store data for each group. Below is the detail description each of the section in the database:

1) Retailer: After registering the shope the retailer can upload product on sale in the table Product. In the Product table, advertised product along with retailer identification will be stored. This identification will be used to advertise the product with market information to the consumer side. Since retailer information is used to track the market for every product, retailers correct registration is necessary before the retailer can start uploading the product.The registration process will have the option of selecting the market and in some cases shop number when registering. In cases brand shopes having shops all over Dhaka city will only need to update their promotional offers. System will automatically display the offer if consumer is nearing a market that has an outlet of that brand. Retailer provides all the necessary information

[6]http://www.androidcentral.com/apps

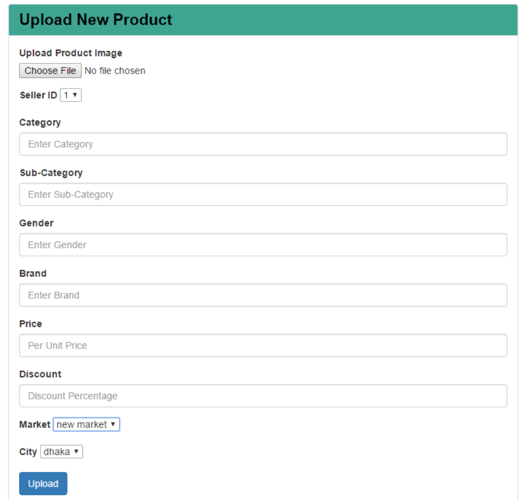


Fig. 2: Promotion Upload form

including shop details, market name, market location, etc in the database. He/She can promote products and can see the popularity of products online. Retailer also can delete product advertisements when the time period is over or anytime retailer wants. Fig 2 shows the promotion upload form for the retailer. On the server, we are using mysql and php. There are 3

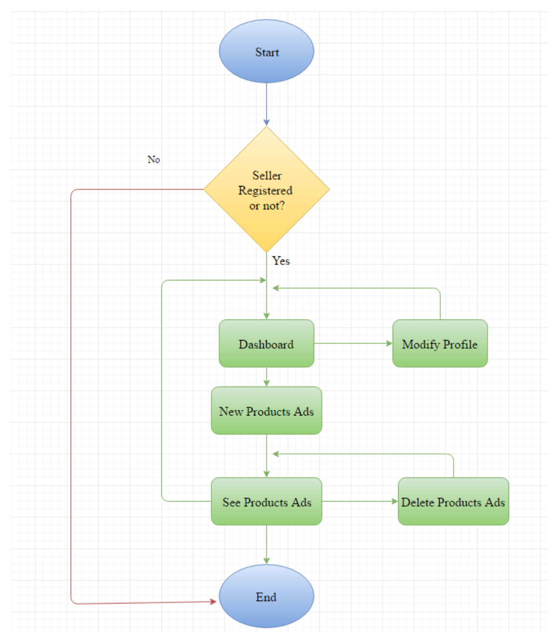


Fig. 3: Retailer Side Flow Diagram

tables for the retailer. First one is retailers information, then market information and promotional offer for products. We have created a php script which will make a json format [7] output of a specific table.

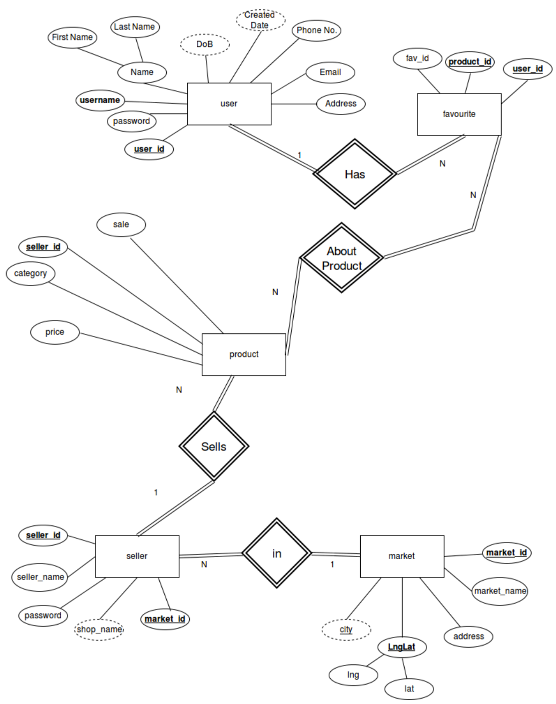


Fig. 4: Database Structure

We are using MySQL in the server-side. There are Five tables in the database. Brief description of the tables are given below. Fig 4 shows table structure.

• User: This is the users table where App user's information will be stored. This table is connected Favourite table
• Favourite: This table will store consumers list of favourite items. Bookmarked product will be displayed to the consumer based on the entry in this table. This table is connected with Products table
• Products: List of products which will be uploaded by the retailer will be stored in this table
• Sellers: Here the retailer's information will be stored
• Market: All markets with the geo location and addresses will be stored in this table

2) Consumer: The objective of this App is to help consumer's shopping experience better. With this goal in mind, the App required that from the consumer side the App will be installed in their mobile phone. As the consumer moves around the city and nearing a market this App will notify any promotion offer in that market. Everyone can see the ads of onsale products and nearest markets. Who Register this App, they will find some extra features. They will have a dashboard where they can see options like Todays Sale, Categories, Nearest Markets,Favourite products, etc. They can bookmark the products. They can see the market position from his/her location in Google Map with the indicate which market has the sale.

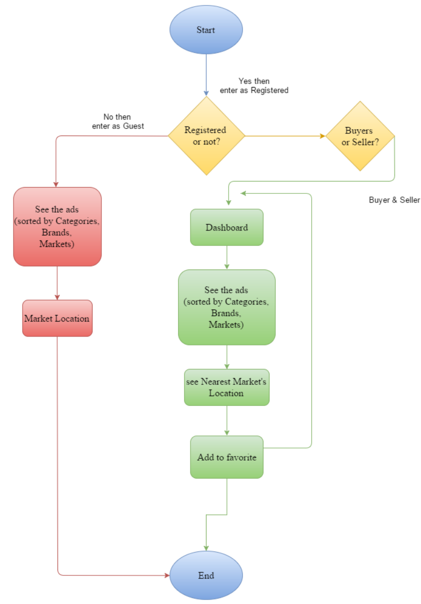


Fig. 5: Consumer Side Flow Diagram

In Registration, Login, Product search and Product showcase, we are using com.loopj.android:android-async-http [9] and org.apache.httpcomponents:httpmime [10] libraies to retrieve the data from server and show the products. We have made some php script to generate JSON format output of the databases data. So we are retrieving the json file using JSON Parsing classes. Fig 5 shows registration process flow diagram.

We are storing the data using SQLite [12] and Sharedpreferences [11] so that user can see this product in offline. But in Offline mode, user can't use the google map and can't get the latest update of the promotions. Fig 6 show the view of the App from consumer side.

Fig. 6: Consumer's View of the Promotions

For the Geo Location, we are using a third party library of a project known as GpsTracker [13]. If the user turn on the GPS and the Internet connectivity, they can use the geo location and see nearest market close to their position. If GPS or the Internet connection isn't available, user can't access the google map. Here we are using Distance API and Duration API which are a part of Google Map for Android v2 [14]. Fig 7 shows what consumer will see when the App starts.

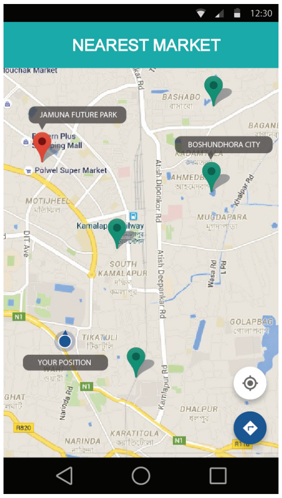


Fig. 7: Consumer's View of the Application

IV. Libraries Used

This section highlights some of the libraries used to develop this App. The main library is Google Play services API which is mainly used for Google map. This API will provide access to location, distance, durations and on in respect of a particular mobile device. Google Cloud Messaging API is used to for push notification to the users about their favorite products. There are some built-in libraries for Material Designs. These libraries are used for better UI design. CircleImageView is a third party library which is used for reshape the images. This is mainly used for displaying images in various size display. GPSTracker is a third party library which is used for find out the users current location and so this app will notify him/her about the nearest on-sale offer. It will take the GPS location in background in every 60 seconds.

V. Challenges and Limitations

This project faces some challenges in the development phase. One of the main challenges was to gather the exact shop/market location. To develop this project we collected exact GPS location manually and inserted those into the database. As protype this is App works only for Dhandmondi area and markers around that area. The developers are currently working on gather GPS data for most of the markets at Dhaka and add those position into the database. Some issues with the user interface design remain which can be solved after collecting user feedback from google play store.

VI. Conclusion

This paper presents a mobile application named CHAAR, that allows consumer to see promotional offers of their favourite products from various outlets. This is designed keeping in mid of a very simple user interface as well as registration process. All the offers will be authentic since only retailers have the option to upload/update a promotional offer. This an App this is developed using the concept of location based service. One of the major point of this App is that this is the first App of this kind in Bangladesh where interest based advertisement is introduced. This is an app designed in such a way that eliminates the security concerns for both the retailer and consumer. Only some basic information is required for registration hence there is any issue regarding user's online security. Authors are working on elimination some of the limitations mentions in the previous section.

References

[1] HP. Bellavista ; A. Kpper ; S. Helal. *Location-Based Services: Back to the Future* IEEE Pervasive Computing, 7(2), pp.85 - 89,2008.

[2] F. Sallabi ; G. Ditsa ; H. El-Khatib ; S. A. Kobaisi. *On-demand Dynamic Location-Based Services Using Web Services* Fifth International Conference on Internet and Web Applications and Services(ICIW), pp.129 - 134,2010.

[3] M. R. Rifat; S. Moutushy; S. Ishtiaque Ahmed;H. Shahid Ferdous*Location based Information System using OpenStreetMap*,IEEE Student Conference on Research and Development (SCOReD), pp.397 - 402,2011.

[4] *Location Based Advertising on mobile*, White Paper, IAB Mobile.

[5] L. Peterson,R. Groot*Location-Based Advertising: The Key to Unlocking the Most Value in the Mobile Advertising and Location-Based Services Markets*, White paper, December 2009.

[6] S. Ahmed, Z.Abdullah,R. Palit, Rokonuzzaman,*A Study of Mobile Application Usage in Bangladesh*,SSRG International Journal of Computer Science and Engineering (SSRG-IJCSE), V. 2(4), Jul.-Aug. 2015.

[7] http://www.json.org/

[8] F. Sallabi ; G. Ditsa ; H. El-Khatib ; S. A. Kobaisi. *Location based advertising framework for mobile and web application developers in Sri Lanka* 2013 International Conference on Advances in ICT for Emerging Regions (ICTer), pp.68 - 74, Dec. 2013.

[9] http://loopj.com/android-async-http/

[10] https://hc.apache.org/httpcomponents-client-ga/httpmime/dependency-info.html

[11] http://developer.android.com/reference/android/content/Shared Preferences.html

[12] https://www.sqlite.org/

[13] http://www.gps-tracker.com/

[14] https://developers.google.com/maps/documentation/android-api/

An Improved Decision based Noise Reduction filter for Salt and Pepper Noise

Golam Moktader Daiyan
Department of Computer Science and Engineering
East Delta University
daiyan@eastdelta.edu.bd

Fahmida Akter
Department of Computer Science and Engineering
East Delta University
fahmida@eastdelta.edu.bd

Abstract

An improved decision based noise reduction filtering approach is presented for the restoration of images that are highly corrupted by salt and pepper noise. It is an enhanced decision based algorithm where noise pixels are detected in several phases based on threshold value. Initially the algorithm will select 3X3 filtering window for detecting corrupted pixels. If all the spatial elements in the window are corrupted, the processing pixel is replaced by previously processed pixel. If the previously processed pixel is 0 or 255 then the algorithm will create a window with a new dimension to define and predict the pixel value. Improved algorithm finds a better result at 9X9 filtering window. It calculates the mean value of all elements in the window and replaces the corrupted pixel. After that, robust estimation algorithm is applied to the proposed filter to remove discontinuity of pixel intensity and smooth the restored image. Experimental result shows that it can provide very high quality restored images, when the noise density is large.

Keywords: *Adaptive Switching Median filter, Salt and Pepper Noise.*

1. Introduction

Images are often corrupted by impulsive noise. There are two models of impulsive noise, namely, salt and pepper noise and random value impulsive noise. Salt and pepper also called as a fixed value impulsive noise because the intensity value of images is changed into 0 or 255 when the image is contaminated by noise. Impulse noise is caused by faulty camera sensors, faults in data acquisition systems and transmission in a noisy channel. Non- linear filtering method i.e. is Median filter are established as a reliable method to remove or reduce salt and pepper without damaging edge details [1, 2]. Several nonlinear filters have been proposed for restoration of images contaminated by salt and pepper noise. Among of them, Standard Median Filter is effective at low noise densities. Several methods have been proposed to remove salt and pepper noise in higher noise densities [3-4]. Computational complexity should consider at the time of implementing a filtering approach. Implementing a filtering with 3X3 mask keeps the computation time minimum. Use of small filtering window for removing noise from the image is insufficient. So that researcher have developed adaptive median filter, where the filtering window size will expanded pixel by pixel to get a noise free pixel. Adaptive Median Filter [5] performs well as low densities. But at high densities the window size has to be increased which may lead to blurring images. After that, researcher has introduced switching or decision technique where noise pixels are detected in several phases based on predefined threshold value. Adaptive switching median filter [6] is one of them; Mainly suffers in three problems which make it inefficient to retrieve image information from high density noise (Experiment shows that not more that 55% noise density) that are losing of local information due to expanding window at a maximum size to get a noise free pixel, increase the runtime due to repeatedly increment of window size and unable to identify noise free pixel when majority of pixels are noisy [7]. Also in the switching scheme the decision is based on predefined threshold value. The major drawback of the switching method is that defining a robust decision is difficult. Noise Adaptive Fuzzy Switching Medan[8] Filter is a hybrid filtering algorithm of Adaptive Median Filter and Fuzzy reasoning based filter where adaptive median filter is used for increasing the window size for getting noise free pixel and fuzzy reasoning is applied to extract local information of the image. The major drawback of this filter is blurring effect. To overcome the above drawbacks, improved decision based noise reduction filtering approach [9, 10] is proposed. Image is initially denoised by applying 3X3 window. If the processing pixel value is 0 or 255 it is processed or else it is left unchanged. If the processing pixel is 0 or 255 then the processing pixel is replaced by median value. Here median value is calculated except considering 0's and 255's. When all the pixels in the 3x3 window are noise then the processing pixel is replaced by last processed pixel if it is noise free. If the last processed pixel is 0 or 255 then the algorithm will create a filtering window with a new dimension. In this stage the algorithm finds a better result at 9X9 filtering window and calculates the mean value of all elements in the window. After, robust estimation algorithm is applied to the proposed filter to remove discontinuity of pixel intensity and smooth the restored image. The proposed filter and several existing method will be tested using several noisy images. The performance of the proposed algorithm will be tested for various levels of noise corruption and will be compared with other existing filters.

The outline of this paper is as follows. Section 2 describes fast adaptive switching algorithm. Section 3 discusses the proposed Algorithm to remove salt and pepper noise. Section 4 deals the illustration of proposed technique. Section 5 deals results and discussions and conclusion is presented in section 6.

2. Review of Fast Adaptive Switching Filter

As the threshold T used in switching median filters varies with different images, it is not easy to set an optimal threshold. An upper threshold T_u and a lower threshold T_l are applied. T_u and T_l are the maximum and minimum pixel values in the filtering window. Therefore, they may vary when processing pixels at different positions. The reason to set T_u and T_l is that the noise pixel is typically the largest or smallest value in the filtering window, especially to the "salt & pepper" noise [5].

The two thresholds are defined as follows:

Tu = maximum{X(i-N, j-N), …, X(i, j), …, X(i+N, j+N)}

Tl = minimum{X(i-N, j-N), …, X(i, j), …, X(i+N, j+N)} (1)

$\alpha(i, j) = 1$ if $x(i, j) = Tu$ or $x(i, j) = Tl$

$\alpha(i, j) = 0$ if $Tl < x(i, j) < Tu$ (2)

$$Y(i,j) = \begin{cases} X(i,j) & \text{if } \alpha(i,j) = 0 \\ M(i,j) & \text{if } \alpha(i,j) = 1 \\ X(i,j) & \text{if } \alpha(i,j) = 1 \text{ and } (2N+1)^2 Nn < (2N+1)^2 \\ X(t1,t2) & \text{if } \alpha(i,j) = 1 \text{ and } Nn = (2N+1)^2 \end{cases} \quad (3)$$

Here,

- X(i’, j’) is a noise-free pixel and its position is the nearest to (i, j).
- If all the pixels in the filtering window are noise pixel, the filtering window is expanded pixel by pixel until a noise-free pixel X(t1, t2) is found.
- Nn represents the number of noise pixels in the filtering window.

3. Proposed Algorithm

Proposed a filtering approach where modification is made on Adaptive Switching Median algorithm (ASM) [5]. During the high density noise, when the all the elements in the processing window at 7X7 mask are noisy the algorithm will create a big processing window with new dimension If the newly selected window contains all 0’s or 255’s or both then the central processing pixel is replace by 0 or 255, which one is more number of times in the selected window. The steps of the proposed algorithm are elucidated as follows.

Algorithm

Step 1: Set Minimum Window size Wmin=3X3 and Maximum Window Size Wmax=7X7. (Set Noisy Image X and Restored Image Z)

Step 2: Read the pixels from the sliding window and store it in S.

Step 3: Compute Smin, Smax, TU, TL, Smed and Np.

Step 4: If Smin<X(i,j)<Smax, where X(i,j) is a processing pixel, then it is considered as uncorrupted pixel and retained. Otherwise go to step 5.

Step 5: If the central pixel X(i,j)=0 || X(i,j)=255 is such That TL<Np<TU then it is considered as uncorrupted Pixel and retained. Otherwise go to step 6.

Step 6: If central pixel X(i,j)=0 || X(i,j)=255 is such that Np<TL then it is considered as corrupted pixel and replace by Smed. Otherwise go to step 7.

Step 7: If Np=TU then increase the window size by to 2 and go to step 2.

Step 8: If W=Wmax then go to step 9.

Step 9: Select the Previously processed pixel Z(i-1,j). If Z(i-1,j-1)~=0 ||Z(i-1,j)~=255 then it is considered as uncorrupted pixel and center pixel replace by previous processed pixel value. Otherwise go to step 10.

Step 10: Create a sliding window with a new dimension. W=Wmax*2 and Compute number of salt (wh) and Pepper (Bl) from the window. If the Wh>Bl the value of center pixel is replace by255 otherwise O.

Here,

- Smin= Minimum intensity value in the window
- Smax= Maximum intensity value in the window
- TU= Number of Pixels at window
- TL=Number of Pixels at window/2
- Smed= Median Value
- Np= Number of Noisy Pixels
- Wh= Number of salt noise
- Bl= Number of pepper noise

The pictorial representation of each case of the proposed algorithm is shown in Fig 1.

The detailed description of proposed algorithm of the flowchart shown in Fig 1 is illustrated through an example in section 4.

4. Illustration of Proposed Algorithm

Each and every pixel of the image is checked for the presence of salt and pepper noise. During processing if a pixel elements lies between “0 and 255”, it is left unchanged. If value is 0 or 255, then it is a noisy pixel and it is substitute by a substitution pixel.

Array labeled with Y displays an image corrupted by salt and pepper noise.

Array labeled Y_1 depicts the current processing window with salt and pepper noise pixels. The square shown in solid line represents the window; and elements inside the circle represent a pepper noise pixel.

Y=

20	189	178	155	199
200	210	213	188	234
168	169	255	255	0
0	255	255	255	0
0	0	255	0	255

Y1=

20	189	178	155	199
200	210	213	188	234
168	169	255	255	0
0	255	255	205	0
0	0	255	0	255

If the current pixel under processing is between 0 and 255, it is left unchanged. Otherwise it will be replaced by a new pixel value estimated using proposed algorithm. For this purpose, the elements inside processing window are arranged as an array and sorted in ascending order except considering " O and 255".

169	188	213	210	205

The element inside the circle is the substitute pixel for pepper noise pixel if the condition: $N_p < W_s \times 0.50$ is satisfied. Otherwise it will remain unchanged.

When the filtering window size is 7×7 and $N_p = Total\ window\ size (7 \times 7)$ then perform the following cases:

Case i): Select the centre pixel of Previous sliding window Z(i-1,j). If Z(i-1,j)~=0 || Z(i-1, j) ~ = 255 then it is considered as uncorrupted pixel and centre pixel replace by previous processed pixel value. The encircled pixel value represents the previously processed pixel.

20	189	178	155	199
200	210	213	188	234
168	169	255	255	0
0	255	255	255	0
0	0	255	0	255

Case ii): Create a window of 15X15 mask and Compute number of salt noise (Wh) and pepper noise (Bl) from the window. If the Wh>Bl the value of centre pixel is replace by 255 otherwise O.

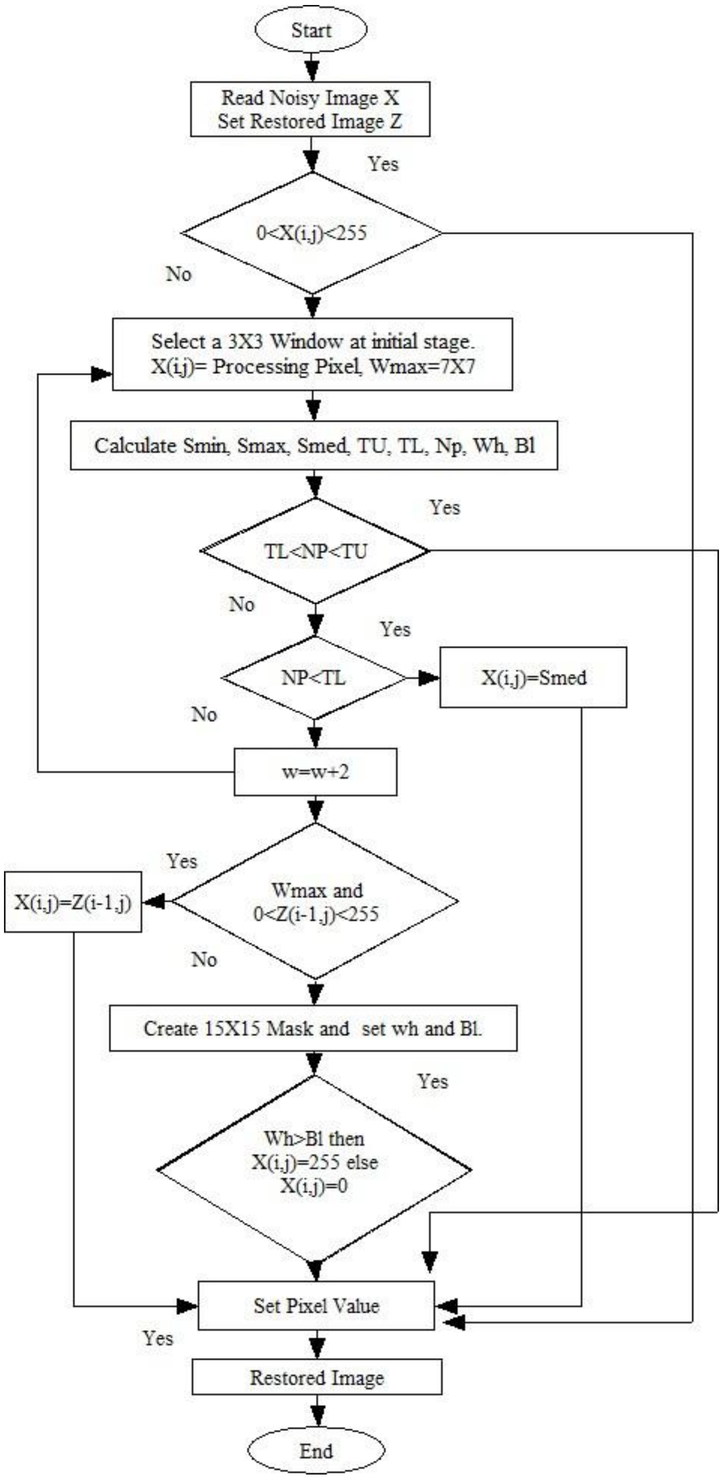
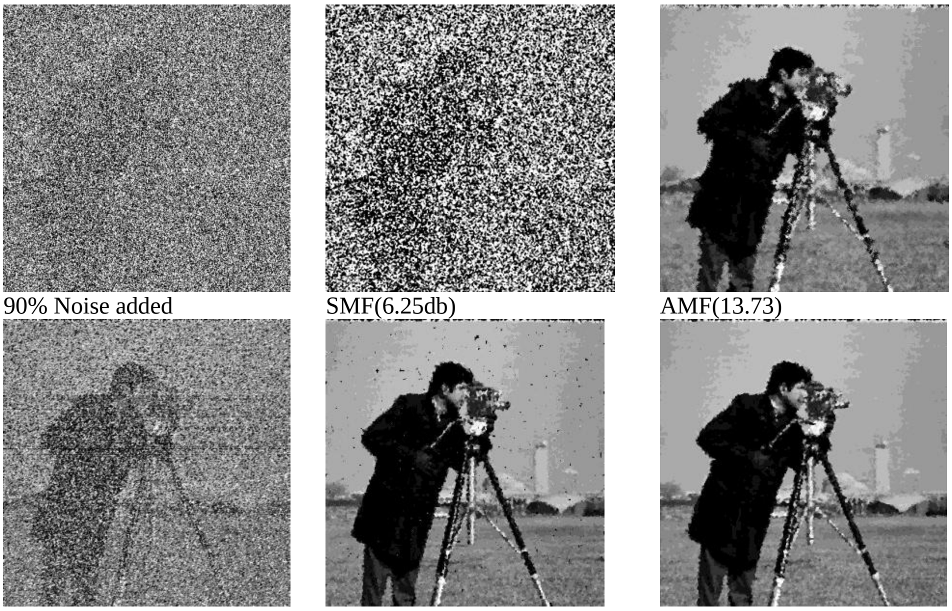


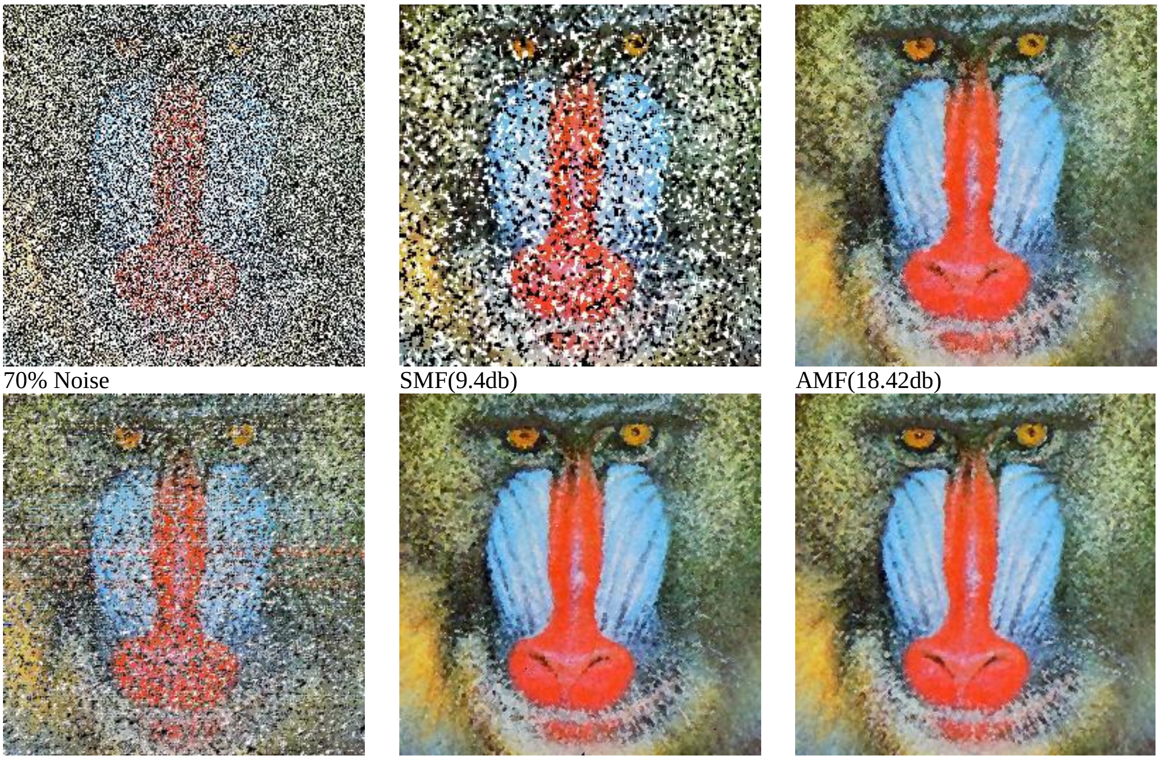
Fig 1: Flowchart of Proposed Algorithm



90% Noise added | SMF(6.25db) | AMF(13.73)

DBA(10.64db) | NAFSM(20.66db) | PA(22.17db)

Fig 2: (from upper left): (a) Cameraman Image corrupted with 90% noise and same image restored with (b) SMF, (c) AMF, (d) DBA, (e) NAFSM Filter and (f) Proposed Algorithm.



70% Noise | SMF(9.4db) | AMF(18.42db)

DBA(13.89db) | NAFSM(19.25db) | PA(20.23db)

Fig 3: (from upper left): (a) Baboon Image corrupted with 70% noise and same image restored with (b) SMF, (c) AMF, (d) NAFSM, (e) DBA Filter and (f) Proposed Algorithm.

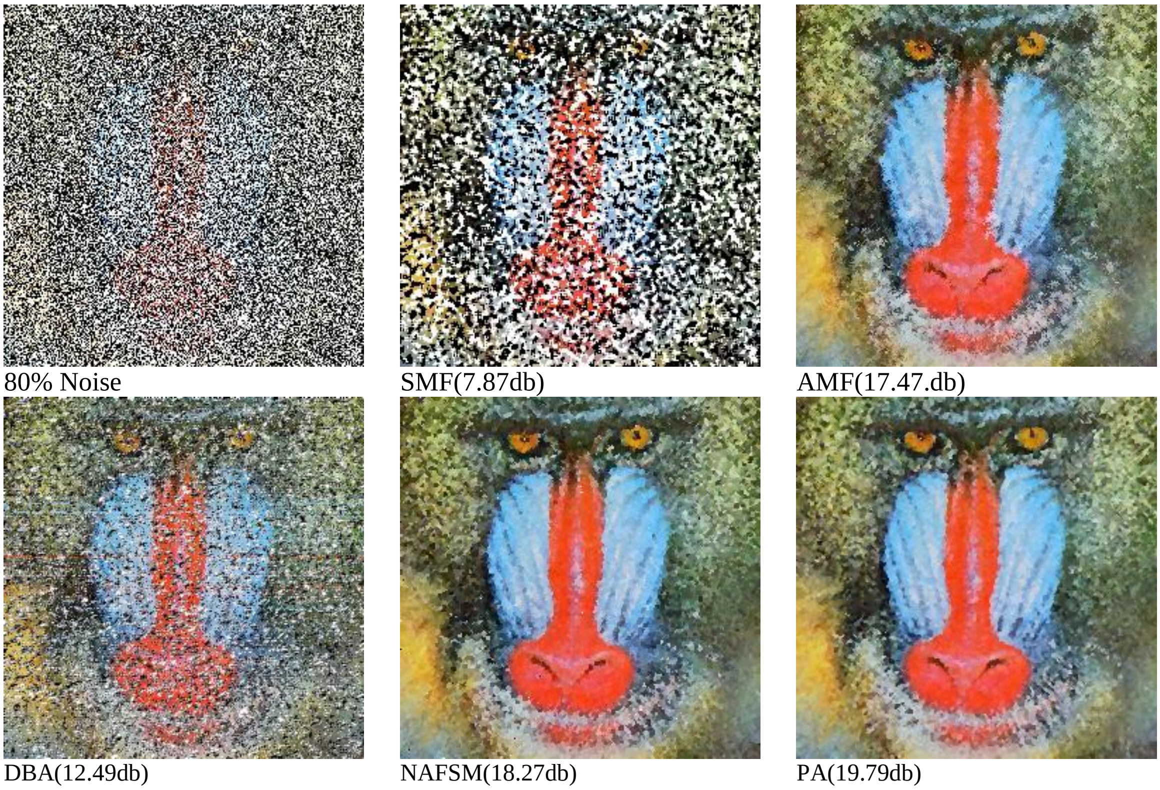


Fig 4: (from upper left): (a) Baboon Image corrupted with 80% noise and same image restored with (b) SMF, (c) AMF, (d) NAFSM, (e) DBA Filter and (f) Proposed Algorithm.

5. Simulation and Performance Analysis

We have used Matlab R2008 as the simulation tool. The proposed filter is tested with different kind of Test images. Images are corrupted by Salt and Pepper noise at various noise densities and performances are quantitatively measured by the Peak-Signal-to-Noise Ratio (PSNR), Mean Square Error (MSE), Image Enhancement Factor (IEF) and Mean Structural Similarity Index (MSSIM) as defined in (1), (2) , (3) and (4) respectively:

$$\text{PSNR}=10\log_{10}\frac{(MAX)^2}{MSE} \qquad (1)$$

$$\text{MSE} = \frac{1}{MN}\sum(Y_{ij}-X_{ij})^2 \qquad (2)$$

$$\text{IEF}=\frac{\sum_{i=1}^{M}\sum_{j=1}^{N}(n_{ij}-r_{ij})^2}{\sum_{i=1}^{M}\sum_{j=1}^{N}(x_{ij}-r_{ij})^2} \qquad (3)$$

$$\text{MSSIM}(X,Y) = \frac{1}{M}\sum_{j=1}^{M}\text{SSIM}(x_j,y_j) \qquad (4)$$

Here M×N is the size of image. X represents the original image, Y represent denoised image.

The noise density is varied from 10% to 80%. The results show improved performance against the existing algorithms at different noise densities for gray scale and colour images. Gray scale image (Cameramen.jpg) and Colour image (Baboon.jpg) are used here to evaluate the performance of the proposed technique.

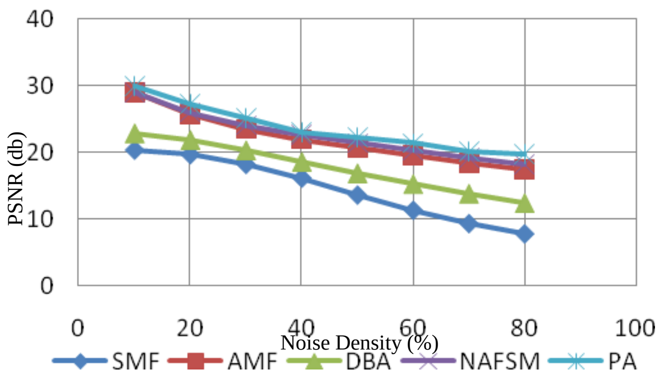


Fig 5: Noise density versus PSNR (db) for Baboon Image.

IEF

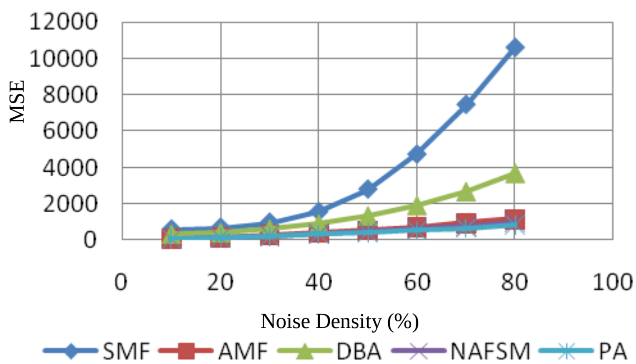


Fig 6: Noise density versus MSE for Baboon Image.

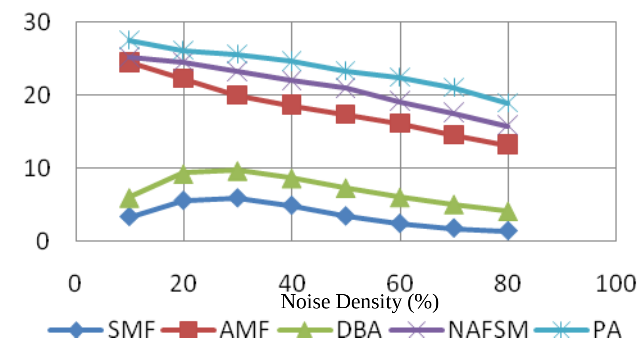


Fig 7. Noise density versus IEF for Baboon Image.

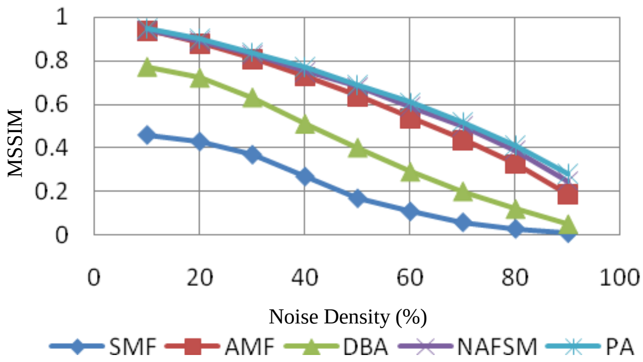


Fig 8. Noise density versus MSSIM for Baboon Image.

The performance of the proposed algorithm for various images at different noise levels is studied. Results are shown in figures 2, 3, 4, 5, 6, 7 and8.

Standard median filter (SMF) replaces the current pixel by its median value irrespective of whether a pixel is corrupted or not. Therefore, the performance is poor. Adaptive median filter (AMF)[3] exhibits improved performance but due to its adaptive nature the computation complexity is much higher. Decision Based Algorithm (DBA) [7] has very good noise removing capacity and good edge preservation at low density noise. Noise Adaptive Fuzzy Switching Median Filter (NAFSM) [9] has improved performance than DBA [7] but its computational complexity is much higher. Figures 5-8 display the quantitative performance of the various algorithms for Lena image. It can be observed that the proposed algorithm removes noise effectively even at higher noise levels and preserves the edges and reduces streaking which is major drawback of DBA [7]. The proposed technique can be good compromise in preference to the adaptive algorithm, DBA [7] and NAFSM [6].

6. Conclusion

An improved decision based noise reduction filtering is proposed in this paper. The algorithm gives better performance in comparison with all discussed algorithm in this paper. Standard Median Filter and Adaptive Median Filter fail to preserve necessary details when noise level more than 30%. Recently proposed Noise Adaptive Fuzzy Switching Median Filtering and Decision Based Algorithm show good de-noising capability at higher density but they produce streaking effect when the noise density is above 80%. The proposed algorithm shows satisfactory results at high noise density and it is designed to preserve necessary details. Due to limited window size it also requires less computation time.

7. References

[1] Castleman Kenneth R, "Digital Image Processing" Prentice Hall, New Jersey 1979.

[2] Rafael C., Gonzalez and Woods Richard E., " Digital Image Processing" 2007.

[3] Hwang H. And Hadded R. A.," Adaptive Median Filter: New algorithms and results," IEEE Trans. Image Process. Vol. 4 no. 4. Pp. 499-502, Apr1995.

[4] Browning D. R.K, "The weighted Median Filter" Communication, ACM, Vol. 27. No.8, 1984, pp 807-818.

[5] Cai. Z. Q, Lee Tracy K. M, "Adaptive Switching Filters", 978-1-4244-4657-5/09, ICICS2009.

[6] Toh Kenny Kal Vin and Isa Nor Ashidi Mat," Noise Adaptive Fuzzy Switching Median Filter", IEEE Signal Processing Letters, Vol 17, No 3, March 2010.

[7] Nair Madhu S., K. Revathy, and Tatavarti Rao, "Removal of Salt-and Pepper Noise in Images: "A New Decision-Based Algorithm", Proceedings of the International Multi Conference of Engineers and Computer Scientists 2008 Vol I IMECS 2008, 19-21 March, 2008, Hong Kong.

[8] Pitas I. and Venetsanopouslos A. N., "Nonlinear Digital Filters: Principles and Applications,'' Norwell, M.A: Kluwer 1990.

[9] Astola J and P. Kuosmanen, "Fundamental of Nonlinear Digital Filtering", CRC Press, 1997.

[10] Hung T.S., Yang G.J. and Tang G. Y. " Fast Two Dimensional Median Filtering Algorithm", IEEE Transaction on Acoustics Speech and Signal Processing, Vol. 1PP13-18, 1979.

[11] Srinivasan E. and Ebenzer D. , " New Nonlinear Filtering Strategies for Eliminating Short and Long Tailed Noise in Images with Edge Preservation Properties", International Journal of Signal Processing", Vol.4. No.3, 1984.

[12] Kong Fanzhi, Ma Wenbin, "A Fast Adaptive Median Filtering Algorithm", 2nd International Conference on Industrial and Information Systems, IEEE 2010.

Cloud Intrusion Detection Model Inspired by Dendritic Cell Mechanism

Azuan Ahmad
Advanced Informatic School
University of Technology Malaysia
54100 Kuala Lumpur, Malaysia
azuan2@live.utm.my

Norbik Bashah Idris
Advanced Informatic School
University of Technology Malaysia
54100 Kuala Lumpur, Malaysia
norbik@utm.my

Mohd Nazri Kama
Advanced Informatic School
University of Technology Malaysia
54100 Kuala Lumpur, Malaysia
mdnazri@utm.my

***Abstract*— Cloud Computing Security is a new implementation of computer technology and opens a new research area and creates a lot of opportunity of exploration. One of the new implementation in Cloud is Intrusion Detection System (IDS).There are problems with the implementation of existing IDS approach in Cloud environment. Implementing traditional IDS need a lot of self-maintenance and did not scale with the customer security requirements. In addition, maintenance of traditional IDS in Cloud Computing system requires expertise and consumes more time where not each Cloud user has. A decentralized traditional IDS approach where being implemented in current Cloud Computing infrastructure will make the IDS management become complicated. Each user's IDS will not be the same in term of type and configurations and each user may have outdated signatures. Inter VM's communication also become a big concern when we implementing Cloud Computing system where communication between Clouds are not monitored and controlled by the traditional IDS implementation. A specific IDS model for Cloud computing is required to solve these problems. In this paper, we develop a prototype of Cloud IDS inspired by Dendritic Cell mechanism.**

Keywords—cloud computing; information security; artificial immune system; intrusion detection; dendritic cell

I. Introduction

With the development of communication networks, Cloud Computing has become critical technology to a modern society. The explosive growth of Internet and Cloud users has motivated the rapid expansion of electronic commerce and other online-based services. Behind convenience and efficiency resulted from these services, there lies a dark side, vulnerability to cyber threats.

As our lifestyle that always depends on network technology or more specifically Internet, we are exposed to at least one of the attacks. In the recent year, even high profile Internet companies like Google, Apple, eBay and Yahoo were hacked by intruders [1]. In Malaysia, estimated losses from electronic hacking are reaching RM 3.3 million within 2012[2]. The increasing importance of computer security motivates various angles of security related research that provide new solutions, which might not be achievable by more conventional security approaches.

Cloud Computing is the new concept of computing where people only need to pay for services and resources without need to place any cost for pysical hardware. With the implementation of Cloud Computing in the application today, it emerges a new technique in software developmnet and deployment. It also change how people are using and managing resources. Cloud Computing can be defined as internet-based computing, where shared resource, software and information are provided to the user on demand [3].

Cloud computing systems are distributed and nesting a lot of resources and private information, therefore because of their nature, cloud computing environments are easy targets for intruders looking for possible vulnerabilities to exploit. When organizations and companies which are using Cloud Computing services, they will move their resource from their own infrastructure to the Cloud infrastructure. If the Cloud is compromised, the organization's resource will be at risk. Cloud Computing systems need protection mechanisms that will monitor the network activity and detect if any intrusion attempts happen within the Cloud Computing infrastructure whether it was from external or internal source [4]. In fact, the cheap availability of significant amounts of computational resources can be regarded as a means for easily perpetrating distributed attacks, as it has recently been observed in several security incidents involving Amazon's EC2 cloud infrastructure.

In addition, there are no specific Intrusion Detection System (IDS) built to protect Cloud Computing systems. Current implementation of IDS in the Cloud Computing systems are still using the traditional way which installing traditional open source or enterprise IDS in the Cloud Computing server to protect the Cloud Computing systems. This traditional IDS implementation, such as on virtual machines (VM), which is considered more vulnerable with diverse security requirements [5]. Implementing traditional IDS need a lot of self-maintenance and did not scale with the customer security requirements. In addition, maintenance of traditional IDS in Cloud Computing system requires expertise and consumes more time where not each Cloud user has [6, 7] [7]. An attack against a cloud computing system can be silent for a network-based IDS deployed in its environment, because node communication is usually encrypted. Attacks can also be invisible to host-based IDSs, because cloud-specific attacks don't necessarily leave traces in a node's operating system, where the host-based IDS reside. In this way, traditional IDSs can't appropriately identify suspicious activities in a cloud environment.

In addition, a decentralized traditional IDS approach where being implemented in current Cloud Computing infrastructure will make the IDS management become complicated. Each Cloud user will install their own IDS and the Cloud Provider will have no authority in managing each Cloud User's IDS. This approach also will affect the services that provided by each Cloud User if the IDS were installed on the host of the Cloud Provider. At the same time, if the IDS was installed traditionally in the Cloud infrastructure and managed traditionally. Each user's IDS will not be the same in term of type and configurations and each user may have outdated signatures. If any attack happens, then each Cloud User's IDS will not treat the event the same way and some of the IDS may not even detect that event. This will bring risk not only to the Cloud User itself but also to the other Cloud Users and in worst case will also affect Cloud Provider and the whole system.

Inter VM's communication also become a big concern when we implementing Cloud Computing system where communication between Clouds are not monitored and controlled. When implementing VM in the system layer of the Cloud, each Guest Operating system (OS) exposed to the risk of being attacked by other Guest OS either intentionally or accidentally. In a way to protect each Cloud, a new method is required to monitor inter VM's activity and detect if any abnormalities occurs and at the same time to block the events from occurring.

II. Related Works

This section reviews the previous work related to Cloud IDS research and prototypes. The number of research focusing in Cloud IDS is increasing rapidly in recent years and there are several solutions that researchers proposed to solve the issues in Cloud Security.

Tupaluka et al. proposed a model based on Virtual Machine Monitor (VMM) or we called Hypervisor within this paper to protect Cloud environment from various attack. This model works on Infrastructure layer of Cloud implementation. VMM have the control over the Cloud resource and this is an efficient way to detect intrusion on the Cloud environment because it has a good visibility of the Cloud resources including network and processing resource of every Cloud host [8].

Gustavo & Miguel in their paper provides a solution to protect Cloud in Software as a Service (SaaS) layer. In their paper, they found that anomaly IDS is a promising technique to be used to protect Cloud application layer [9] .

Viera et al. proposed a Grid and Cloud Computing Intrusion Detection System (GCCIDS). Their prototype used Artificial Neural Network (ANN) as the machine learning algorithm to train the system and developing their prototype using Grid-M middleware. They proved that their system had low processing cost while maintaining satisfactory performance for real-time implementation, since it performed the analysis individually on each node, resulting in lower data exchanges between nodes, thus decreasing the complexity of the system [10].

III. Dendritic Cell Mechanism

This section describes Dendritic Cell process and their components.

A. *Dendritic Cell Inspiration*

Dendritic cells (DC) are the main function in natural immune system by which the innate immune system collects and present antigens to the adaptive immune system for processing. Dendritic cell exist within three states immature, semi-mature and mature dendritic cell where immature dendritic cells are reside in tissues throughout the body for collecting antigens and signals for processing, semi-mature dendritic cells is the results from immature dendritic cells that collect antigen and signal in a environment that have safe signal more than danger signal and mature dendritic cell on the other hand is the results from immature dendritic cells that collect antigen and signal in a environment that have danger signals more than safe signals. Dendritic cells are especially abundant in tissues where pathogens may enter body, such as skin, lung and gastrointestinal tract. Figure 1 simulate the Dendritic Cell Maturation Process by stimulation of various signal

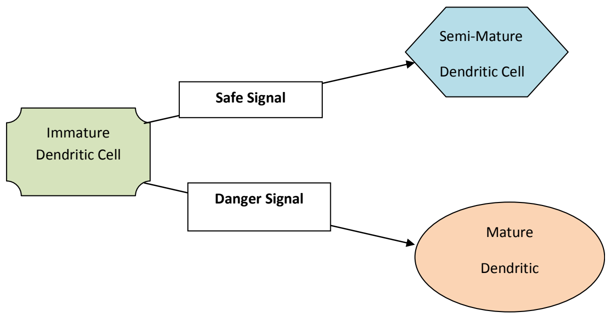


Fig. 1. Dendritic Cell Maturation Process

B. *Antigen*

Dendritic cells ingest nearby pathogens and cellular debris and process this ingested material and use molecular structures on their surfaces to present any antigen found. Dendritic cells also bind with signalling molecules that affect their functioning and provide stimulus for maturation.

As they mature, dendritic cells leave the peripheral tissues and migrate to the lymph nodes and other lymphatic organs. In the paracortex of lymph node, a dendritic cell interacts with lymphocytes, such as T-cells presenting antigens for further processing by the adaptive immune system.

C. *Signals*

The Danger Model holds that the maturation of dendritic cells is controlled by signaling molecules named Pathogen

Associated Molecular Pattern (PAMP), danger, safe and inflammation signals found in the surrounding tissue. Tissues experiencing stress or damage emit danger signals while healthy, unstressed tissues emit safe signals. Some molecular patterns commonly found along with bacteria and other pathogens also act as danger signals.

Sufficient stimulus by danger signals causes dendritic cells to become fully mature. This causes them to express signalling molecules that indicate the antigens they present were found in a dangerous environment. Mature dendritic cells promote immune reactions to the antigens by the adaptive immune system. On the other hand, sufficient stimulus by safe signals causes the dendritic cells to become semi-mature. Semi - mature dendritic cells indicate that their antigens were collected in a safe environment and tend to suppress immune response to these antigens.

PAMPs are molecules produced my microorganism. These molecules are not unique to pathogens but are produced by microbes. PAMP molecules are an indicator to human immune system that a non-host entity was presented. Specific PAMPs bind to specific receptors on dendritic cells which can lead to production of both co-stimulatory molecules and interleukin-12 (IL-12) which related to danger signal. In our immune system, PAMPs is as a biological signature of abnormality. In this Cloud IDS model, PAMP is interpreted as a signal which is an indicator of an abnormality. This is presented by the detection of intrusion based on detection signature.

Danger signals are the signals release when a necrosis happens in the tissue cells. Necrosis is the unexpected or forced death of tissue cell that indicate something abnormal was happened in the tissue. The release of danger signal is the indicator of damage to the tissue against which the immune system is trying to protect. The sufficient exposure to the danger signal causes DC maturation to the fully mature state. Potency of danger signal is less than PAPMs, meaning that a higher concentration of danger signal molecules are needed in order to produce a response of the same magnitude as similar concentration of PAMPs. Concentration is the number of molecules of signal per unit volume. Within this thesis, danger signals are indicators of abnormality but have lower value of confidence that the PAMP signal. Danger signals expression is an indication that antigen in a dangerous context thus lead to the activation of the adaptive immune system.

Safe Signals is the signals release as a result of healthy tissue cell functions normally released during a normal cell death or known as apoptosis in medical term. A molecule named Interleukin-10 (IL-10) is produced as a result of the presence of safe signal in the tissue. The production of IL-10 indicates that antigen collected by DC was found in a normal, healthy tissue thus will suppress the immune reaction to the antigen. When a tissue contains cells undergoing both apoptosis and necrosis, the receipt of safe signal will suppress the production of IL-12 molecules is response to the danger and PAMP signals present in the tissue. This is one of the mechanisms in the immune system to prevent false positives.

The presence of inflammatory signals in human tissue is insufficient to initiate maturation of an immature DC. However, the presence of inflammation not only implies the presence of inflammatory cytokines but also the temperature increased in the affected tissue. The rates of reaction also increased because of the increasing heat and inflammatory cytokines initiate the process of dilating blood vessels result in increased number of cells to the tissue under distress. In this model, inflammation has the effect of amplifying the other three categories of input signals. The result is an increase in the artificial DC's output signals. An increase in inflammation implies that the rate of DC migration will increase, as the magnitude of the CSMs produced by the DC will occur over a shorter duration and hence resulting in a shortened DC life span in the tissue compartment. The presence of inflammatory signals alone is insufficient to instruct the immune system how to behave accordingly.

Within Cloud IDS model, input signals that indicate normal activity are known as safe signal. This signal is interpreted as data which indicates normal system or data behaviour and high level of this signal will increase the output signal value for the semi-mature DC. The receipt of a high safe signal will reduce the cumulative value of the mature DC.

IV. Algorithm

This section explains the algorithm of DC mechanism for Cloud IDS model. Inspired by the activities of DC in human tissue, Cloud IDS model try to mimic the same process as a solution in protecting Cloud network from intrusions.

Depicted in Figure 3, each monitored Cloud network activity is viewed as Antigen and the Internet Protocol (IP) address of each packet is taken as the Antigen identity. The Cloud IDS perform multiple signal and antigen sampling. Cloud IDS model will collect three signals from the Cloud environment; PAMP, Safe Signal and Danger signals linked to a specific antigen that trigger that signals. The signals then

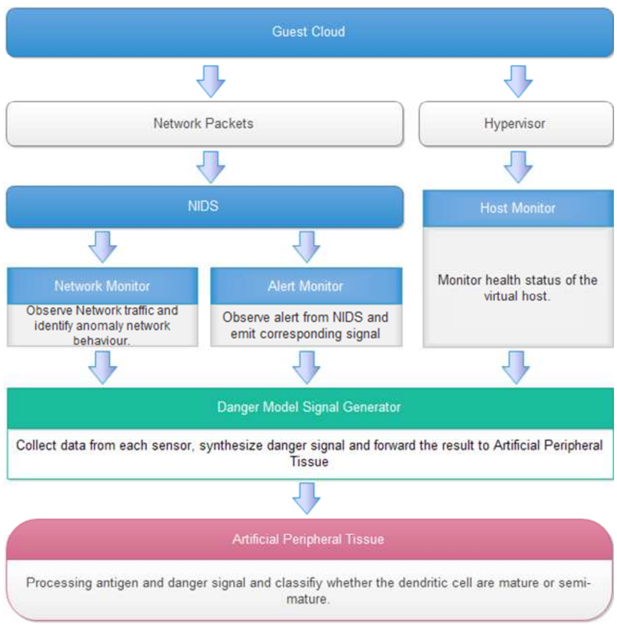


Fig. 2 Cloud IDS model

will be cumulatively group based on the DC. In our experiment, we consider each DC handles a specific antigen. Based on the collected input signals, the DC will be transform into either three outputs states; co-stimulatory signal (CSM), semi-mature and mature. When the DC exceeds the maturation threshold, in our case the monitoring time limit, the DC stop monitoring and the output signal values will be analysed. When learning ends, antigens appear in different contexts. In the last step, the potential anomalous antigen is determined based on the collected context known as the mature context antigen value (MCAV), the anomalous antigen is determined as:

$$MCAV = \sum mature / (\sum mature + \sum semi\text{-}mature) \qquad (1)$$

The antigens with a greater than the anomaly threshold are classified into the anomalous group while the opposite are considered as the normal category.

V. CLOUD IDS MODEL

This section describes Cloud IDS model, a model for detecting any threat and intrusion attempt for Cloud environment. Cloud IDS model imitate the functionality of dendritic cells in human immune system that protect our body from infection of pathogen and bacteria.

The Cloud IDS Model draws inspiration from the dendritic cell maturation process of the natural immune system. Figure 2 depicts an overview of the Cloud IDS model. This model synthesizes antigens from packets observe on the cloud network. It also synthesizes danger model signals from observed events and the state of the network and guest cloud. This model then classifies antigens as dangerous or safe and provides this information in detecting any threat to the Cloud environment.

The Cloud IDS model emulates and make use of the functions and activity of the dendritic cells in the body tissue of the HIS and applying the concept in protecting Cloud environment. This model consists of a set of danger model signal generator, a misuse-based network intrusion detection system (NIDS) and artificial peripheral tissue (APT) where the dendritic cells, antigens and danger model signals interact.

Emulating the activity of dendritic cell required this model to have two most important elements of immune system, antigen and signals. Cloud IDS model captures and decomposes network packets collected from the private cloud environment and at the same time, antigens will be extracted from the network packets by the selected features of the network packets. This model also synthesizes danger model signals from external data sources.

This model provides two types of output, the sequence of alerts from the misuse-based NIDS and a sequence of artificial dendritic cells which presenting processed antigens and their corresponding dangerous or safe context. The dendritic cells then are process in the APT to get the maturation level of each cloud area. Figure 3 presents the elements of Cloud IDS model and the flow of data in this model.

A. *Antigens and Signals Representation*

Cloud IDS model uses two primary source of information as a primary data; antigens and signals. The antigen represents the cloud network traffic, which each monitored network packet resulting in the synthesis of a corresponding antigen.

Cloud IDS model contains two types of feature; address and protocol features. Address features are 32-bit, unsigned integer value Internet Protocol version 4 (IPv4) address found in the packet header information. Example of address features are 192.168.0.1 and 172.16.112.20. IPv4 address is used as the source or destination address for every network communications. On the other hand, protocol features are 32-bit, unsigned integer derived from the protocol value found in the IPv4 packet header information. There are two commonly used protocols available Transmission Control Protocol (TCP) and User Datagram Protocol (UDP) and for each protocol, port number is assign to represent what service they provide. Port number is the least-significant 16-bits of the feature in the packet header information, bits 0 to 15.

Cloud IDS model collects signals from the Cloud network by implementing signal sensors on each Cloud node. Each Cloud node consists of three signal sensors; host monitor, alert monitor, and network monitor. Each signal includes two functional elements. The first element is the antigen feature value where this indicates the antigen that produces the signals. The second element is the signal level. This is an integer value that determines the degree of danger or safety each corresponding signal represents. A signal with high level of

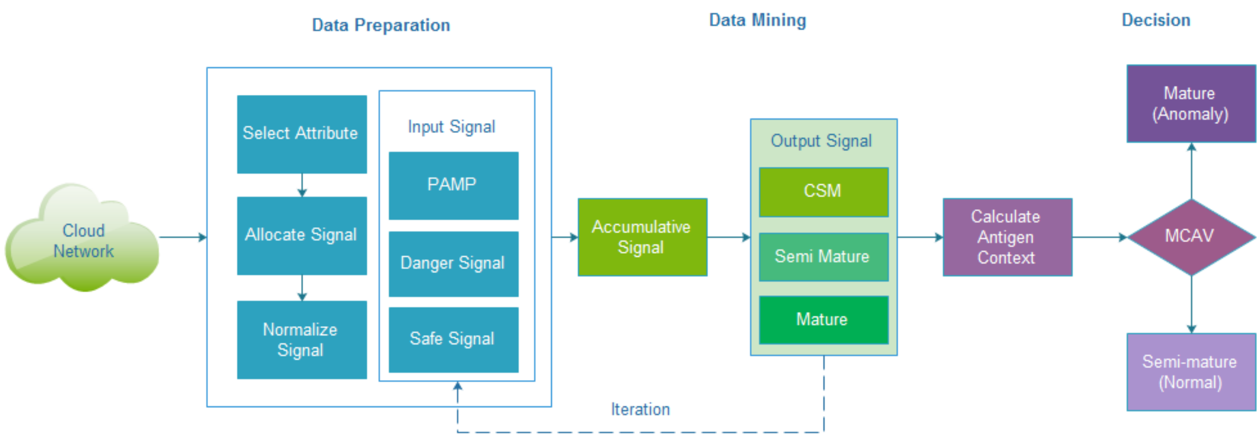


Fig. 3. Algorithm for Cloud IDS model based on Dendritic Cell mechanism

danger signals indicate a signal was collected in an area of danger and on the other hand, a signal with high level of safe signal indicate that the signal was collected in a safe area.

B. *Host Monitor*

Host monitor observe the state of each guest cloud host and emits signal based on the health status for each of the Cloud user. The state of the cloud host will affect the immune response in Cloud IDS similar as in HIS where the tissue states affect the response of the human immune system. Any host state that showing the damage on the host promotes immune reactions while healthy host state suppresses immune reaction.

Host monitor continuously monitor the status of each cloud host in the cloud environment by using multi-agent system. When activated, host monitor gather monitoring data using agent that installed on each Cloud host through secured channel. Figure 4.1 describe the Host Monitor activity in monitoring each Cloud nodes.

C. *Network Monitor*

Network Monitor observes and analyse network traffic in the cloud network and emitting danger signal based on the state of the network traffic. Network Monitor identify anomaly in the cloud network by comparing the current network behaviour with the normal traffic behaviour or known as normal traffic profile. Network behaviour that significantly different from the normal traffic profile are an indication of anomaly and result in emission of danger signals. Observations that similar or within the normal traffic profile is considered normal and result in emission of safe signals.

Network monitor assist Cloud IDS in detecting any anomalous cloud network traffic that may indicate a threat to the cloud environment by emitting or suppressing related signals. This process is analogous to the effect of tissue stress on the HIS. Tissue under stress emits chemical signals that promote immune response while unstressed tissue suppresses immune reactions.

D. *Alert Monitor*

Alert monitor analyse the alerts emitted by the NIDS and generate a corresponding danger signal. This will results evidence of danger seen in the network packets to affect the immune response. This is inspired by the ability of dendritic cells in detecting the presence of pathogens through reaction to Pathogen Associated Molecular Pattern (PAMP) signals collected in body tissue.

VI. CONCLUSION

In this paper, IDS methodology was combined with Dendritic Cell Algorithm mechanism to provide a solution in detecting any attack targeting the Cloud environment. Cloud IDS model mimics the activity and process of Dendritic Cell which is known for detecting and killing any pathogens that infected human tissue and cells. The successful of Dendritic Cell in protecting human body will also bring a success in protecting Cloud environment if the same mechanism are being implemented in the real world applications.

References

[1] BBC, "Google and Apple among hundreds hit in high-profile Pakistan hack," 26th November, 2012 2012.

[2] CyberSecurity. (2012, 19th November 2012). *Hacking costs Malaysia MYR 3.3 million*. Available: http://www.cybersecurity.my/en/knowledge_bank/news/2012/main/detail/2249/index.html

[3] P. Mell and T. Grance, "The NIST definition of cloud computing (draft)," *NIST special publication*, vol. 800, p. 145, 2011.

[4] D. Gollmann, "Computer security," *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 2, pp. 544-554, 2010.

[5] S. Subashini and V. Kavitha, "A survey on security issues in service delivery models of cloud computing," *Journal of Network and Computer Applications*, vol. 34, pp. 1-11, 2011.

[6] I. Cymtec Systems. (2012). *Scout Cloud-Enabled IDS Fact Sheet*. Available: http://go.pardot.com/l/12332/2012-04-16/kjv2/12332/9341/Cymtec_Scout_IDS_Fact_Sheet_021412v2.pdf

[7] W. Yassin, N. Udzir, Z. Muda, A. Abdullah, and M. Abdullah, "A Cloud-based Intrusion Detection Service framework," in *Cyber Security, Cyber Warfare and Digital Forensic (CyberSec), 2012 International Conference on*, 2012, pp. 213-218.

[8] U. Tupakula, V. Varadharajan, and N. Akku, "Intrusion detection techniques for infrastructure as a service cloud," in *Dependable, Autonomic and Secure Computing (DASC), 2011 IEEE Ninth International Conference on*, 2011, pp. 744-751.

[9] G. Nascimento and M. Correia, "Anomaly-based intrusion detection in software as a service," in *Dependable Systems and Networks Workshops (DSN-W), 2011 IEEE/IFIP 41st International Conference on*, 2011, pp. 19-24.

[10] K. Vieira, A. Schulter, C. Westphall, and C. Westphall, "Intrusion detection for grid and cloud computing," *IT Professional Magazine*, vol. 12, p. 38, 2010.

A Routing Protocol for Cognitive Radio Ad hoc Network

I. Introduction

Due to the expansion of wireless communications, it is almost impossible to cope with allocation of spectrum as it is limited and finite resource. Therefore, it is vital to ensure the best usage of this limited resource to meet the recent demand of radio frequency. The traditional spectrum allocation methods that are authorized by different regulators are not intelligently handled. It is observed that the licensed spectrum remains unoccupied most of the time [1] where the regulatory body is unable to allocate the frequency for upcoming new wireless devices [2]. This temporarily unused portions in the licensed spectrum are called spectrum holes [3]. Spectrum hole is define as the frequency band that has been allocated to a licensed user but it is unutilized at a particular time or location. To utilize these unoccupied band, resolve the spectrum scarcity for new wireless application and make spectrum allocation more dynamic, cognitive radio (CR) technology is introduced [4]. This intelligent technology has introduced opportunistic access of unlicensed user, known as secondary user (SU), to use licensed band without interfering the licensed user, known as primary user (PU) [5].

Cognitive radio networks (CRNs) can be classified as infrastructure-based CRN and cognitive radio ad hoc networks (CRAHNs) [5]. The infrastructure-based CRN has a central network entity to manage the network such as base-station (BS) in cellular networks. Instead, CRAHN does not have any infrastructure backbone and each CR user can communicate with other CR users through ad hoc connection [6]. Most of the work in the field of CRAHN has focused on channel scarcity problem at the lower layer (PHY, MAC), while routing in CRAHN is largely unexplored. Routing in CRAHN is very important task that have a great effect on the overall performance of the network. Routing in CRAHN differs with routing in traditional ad hoc network as it has to adopt to dynamic changes of spectrum due to stochastic behavior of PU and SU. Moreover, routing protocol in CRAHN must deal with heterogeneity of resources (available channel and available energy).

In this paper, we designed an efficient clustering mechanism and on top of that we proposed a robust routing algorithm to ensure the certain level of QoS in the network. A novel spectrum-aware clustering algorithm is introduced that jointly considers the available spectrum and the power level of the nodes to form the clusters. The proposed clustering algorithm enables the network to be more robust to PUs' activities as well as node mobility. The proposed routing protocol is a hybrid protocol that uses a proactive method for intra-cluster routing and a reactive method for inter-cluster routing. The path selection in proposed protocol is defined as a multi-objective optimization problem, where proposed approach select the path with the low density. The idea is that packet will face long delay due to collision is highly dense area.

II. System model

Fig. 1 shows considered system architecture, in which we assume that there are N_{SU} cognitive radios and N_{PU} primary users deployed in CRSN. A given number of non-overlapping orthogonal channels $\{Ch|Ch_i,\ i = 1, 2, \ldots n\}$ are available, and each channel has a unique channel ID. Each node is aware of its location and each node has a single half-duplex cognitive radio transceiver, which is capable of detecting and utilizing spectrum holes in a distributed and efficient way. Cognitive radios or SUs coexist with PUs and opportunistically and conditionally access the channels.

A discrete-time Markov chain is employed to PUs' channel-usage patterns [7], which means that the PU's may change their state (i.e., channel usage) after each process or step. Channel availability for each node is related to the node's physical location. Similar to IEEE 802.22 standard, SUs use an available channel only when it is not occupied by PUs. By detecting PU's presence, SU vacates the channel. The overlay spectrum sharing model [8], is used in our network as simplified interference avoidance model.

A dedicated control channel has been considered in the network. Nodes forming the cluster become cluster heads (CHs), which are responsible for inter-cluster communication as well as intra-cluster channel access control. Inter-cluster communication is relayed by gateway nodes (GW). GWs are the nodes which are in the border of two neighboring clusters and can hear both cluster beacons as shown in Fig. 1. GW inform CH about its status through control channel.

III. Cluster formation

The proposed clustering mechanism divides the network into clusters based on three values: spectrum availability, node power level and node current speed. In the proposed clustering scheme, clusters are formed with the neighboring nodes in an ad hoc topology.

Nodes in the proposed architecture exchange their available channels list ($ACL_i(t)$) based on the spectrum sensing information. Each node generates its own neighbor list N_i where i=1, 2, 3,. . . , n using neighbor discovery mechanism. Afterwards, the cluster formation phase starts. Cluster formation is defined

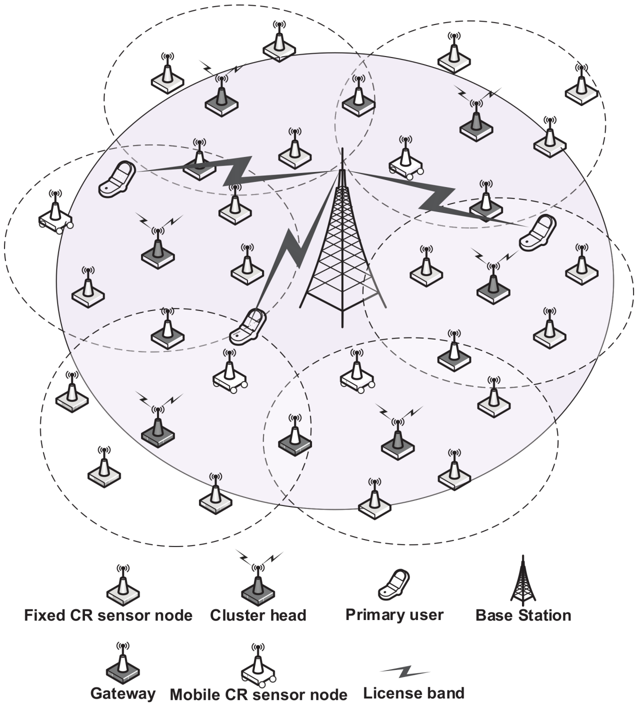


Fig. 1: System architecture.

as a maximum vertex biclique problem [9] to include more nodes inside each cluster while making sure of existing enough number of idle channels for intra-cluster communication.

At first, based on neighbor list, N_i and $ACL_i(t)$, every CR_i creates an undirected bipartite graph $G_i(A_i, B_i, E_i)$. Graph G (V, E) is called bipartite if vertices set V can be split into two disjoint sets A and B where $A \bigcup B = V$, such that all edges in E connect vertices from A to B. Here, $A_i = CR_i \bigcup N_i$, and $B_i = C_i$ where, C_i is the $ACL_i(t)$ of CR_i. An edge (x, y) exists between vertices $x \in A_i$ and $y \in B_i$ if $y \in C_i$, i.e., channel *y* is in the available channels of CR_i.

To choose an optimal CHs among all nodes, we define a parameter named a cluster head election value (CHEV). In this paper we formulate the choosing of a CH as a maximization problem, which can be defined as follow:

$$i_j{}^* = \max_{i_j}(CHEV_{i_j}), \quad 1 \leq i_j \leq CM_j, \tag{1}$$

$$CHEV_{i_j} \quad \propto \quad N_{i_j}{}^{Ch_{i_j}}, \tag{2}$$

where i_j indicates the node *i* in cluster *j*, N_{i_j} is the total number of neighboring nodes to node *i* in cluster *j*, and Ch_{i_j} is the total number of common channels for node *i* in cluster *j*.

The idea behind choosing the *CHEV* value as it is given in (1) and (2) is to choose the node with the highest number of common channels and the highest number of neighbors to be the cluster head. This gets the cluster more flexible to the PU appearance and spectrum mobility as well as avoiding having a high number of clusters in the network.

Since CH is responsible about cluster stability, it should be the most powerful among the cluster nodes. Also, for producing a high mobility aware MAC protocol by avoiding frequent reclustering due to CH movement, the CH should be the lowest speed node among the cluster nodes in the network. By combining these two important features of CHs, we define the constant of the relationship given in (2) as follows:

$$CHEV_{i_j} = W_{i_j} \times N_{i_j}{}^{Ch_{i_j}}, \tag{3}$$

$$W_{i_j} = \frac{\gamma_{i_j}}{\sum_{i_j}\gamma_{i_j}} \quad where \quad \gamma_{i_j} = \frac{E_{i_j}{}^{\alpha}}{V_{i_j}{}^{\beta}}, \quad \alpha, \beta \in R+, \tag{4}$$

where W_{i_j} is a normalization factor that indicates how well node i_j is powerful and static in relation to the other nodes in the cluster. γ_{i_j}, which is always a positive value, is the proposed parameter to indicate the relationship between the node energy and speed. Meanwhile, α and β are design parameters for prioritising the speed and energy based on the application requirement.

To avoid having a large *CHEV* value, we take the log of *CHEV* as a final selection metric of CHs. Thus, the maximization problem in (1) can be written as follows:

$$i_j{}^* = \max_{i_j}(\log(W_{i_j} \times N_{i_j}{}^{Ch_{i_j}})). \tag{5}$$

This maximization problem can be simply solved by the well known descending sorting algorithm [10]. Therefore, a node with the highest log(*CHEV*) value forms the cluster and becomes the cluster head (CH). For example, when node *a* and *d* have 4 shared channels, while node *d* has a higher number of neighbors compared to node *a* under the conditions that they are static and have the same amount of power, node *d* will be the CH. If the log(*CHEV*) value of a node CR_i is smaller than its neighbor, CR_i joins the neighbor, which has the highest log(*CHEV*) value, as cluster member (CM). Once the clusters are formed, CHs prioritize other cluster members based on log(*CHEV*) for the reserved cluster head (RCH) selection. CM with the highest log(*CHEV*) becomes the RCH for the cluster. The RCH takes charge of the cluster if current CH moves out, which reduce the possibility of re-clustering.

IV. Routing

By using the clustering mechanism, the proposed routing protocol uses both a proactive and a reactive routing in an adjustable hybrid manner. Each node is aware of the topology of the cluster that it belongs to. Therefore, the packet transmission inside each cluster occurs in a proactive manner. However, no matter how big is the network size, nodes does not require to know the topology of the whole network and the updates are propagated only locally inside each cluster. Meanwhile, routing between clusters happens in a reactive manner using CHs and gateways.

Source node sends the route request (RReq) to its cluster head. If the destination node is in the same cluster then cluster head informs the source node about the route to the destination. Otherwise, CH broadcasts the route request (CHRReq) to the adjacent clusters using the GW nodes. Each CH that receives the CHRReq packet, it checks that whether the destination node is within its cluster members or not. If a CH finds the destination node is within its members, it replies the CHRReq by CHRRep. If it does not find the destination among its members, it adds its id, hop-count, cluster channel and cluster routing weight (CRW) to the packet and forwards the message to the adjacent clusters. Discussion about the function of CRW will be given in the following section.

Since in the proposed routing mechanism, routing queries propagate only among the CHs, thus it relatively uses small number of query messages.

A. Path optimization

Considering that improvement of different routing metrics may result in different routing paths, it is necessary to combine different individual routing metrics to form a global metric to achieve a performance trade-off among different routing metrics. In this paper, as we consider to improve multiple factors in routing such as delay, bandwidth and hop-count, path optimization is defined as multi-objective optimization as follows:

$$\begin{aligned} \min_x F(x) = (f_1(x), \quad f_2(x), ..., \quad f_n(x))^T \\ s.t. \quad n \geq 2, \quad f_i(x) \to \mathbb{R}^+ \end{aligned} \tag{6}$$

where n is the number of objective functions, $F(x)$ is a vector of objective functions $f_i(x)$. It is needed to be noted that there is no single global solution to multi-objective optimization problems and it is more of a concept than a definition. To solve the above multi-objective optimization, we use very common method called weighted sum method [11].

$$M(r) = \sum_{1}^{n} w_i f_i(x) \tag{7}$$

where $M(r)$ corresponds to the global objective function F and r defines the chosen route. $W \triangleq w_i$, $w_i > 0$, $1 \leq i \leq n$ and $\sum_{i=1}^{n} w_i = 1$ is a weight vector defined by each node and reflect the relative importance of that particular object.

In this paper, we aim to minimize the total delay in routing which is the main objective. The delay at each node is defined as follows:

$$D_{total} = \quad D_t + D_s + D_b + D_q \tag{8}$$

where D_t is transmission delay and is defined as the time taken by each node to transfer the packet to the next hope. D_s is the switching delay and defined as the delay time caused by frequency band switching. Since in this paper we assume that the clustering scheme and the intra-cluster nodes are using the same frequency band, we ignore D_s and consider it to be zero. D_b is the backoff delay time that occurs when carrier sense fails due to collision or detecting another transmission. D_q is the queueing delay time that highly depends on the traffic load, and it is defined as the total time a packet spends in a queue before transmitting.

Suppose we have N hops between source to destination, so the backoff delay at hop n denoted by $D_{b,n}$ and queueing delay is denoted by $D_{q,n}$. If the packet length is fixed, transmission delay would be fixed in all hops and it is denoted as D_t as before. Therefore, the total delay over N hop is

$$D(N) = \sum_{n=1}^{N} (D_t + D_{b,n} + D_{q,n}) \tag{9}$$

And average delay over N hop is

$$E[D(N)] = N(D_t + \overline{D_b} + \overline{D_q}) \tag{10}$$

Equation 10 shows that multihop delay linearly increases with the increasing number of hops.

Higher bandwidth assures faster transmission, where values of D_q and D_t are low. Meanwhile, chances of collision are increased in a highly dense network, where increased chance for collision increases D_q. This is because the packet should remain in the queue longer till the node find available free channel to transmit. Also it increases the D_b, as nodes have to perform backoff procedure more often due to collision. Therefore, D_q and D_b have direct relationship with node density. We consider node density on any area as the number of neighbors that the node has. Higher number of neighbors means higher node density in that area. Since all traffic flow through CHs, we only consider node density around CHs where number of neighbors for each CH represents node density.

$$\begin{aligned} D_q, \quad D_b \quad \propto \quad Node \quad density \\ D_q, \quad D_t \quad \propto \quad \frac{1}{bandwidth} \end{aligned}$$

According to [12], equation 11 is used for calculating the 1-hop network density:

$$\mu(r) = \frac{N\pi r^2}{A} \tag{11}$$

where *N* is the total number of nodes in the network, *r* is the transmission range of each node and *A* is the size of the area where the network is deployed. Bandwidth of cluster is calculated similar to [13]. Defining $R_s(t)$ as the achievable data rate while PU correctly detected as idle without any false alarm, and $R_f(t)$ as achievable data rate during the falsely sensed idle channel, we have:

$$R_s(t) = (P_{off} - P_f)\frac{T - \overline{T}_s - \overline{T}_o}{T}\beta \log_2(1 + \frac{S_r^s}{n(t)}) \tag{12}$$

$$R_f(t) = (P_{on} - P_d)\frac{T - \overline{T}_s - \overline{T}_o}{T}\beta \log_2(1 + \frac{S_r^s}{n(t) + S_r^p}) \tag{13}$$

where P_{on} and P_{off} are defined as the probability of a channel being in occupied state and channel being in idle state, respectively. P_d is defined as probability of detection and P_f

as false alarm. T is the SU maximum frame period, while T_s is defined as the sensing period and and T_o is defined as overhead of negotiating the traffic channel between the pair of transmitter and receiver. S_r^S is the received signal of PU by SU during spectrum detection. S_r^p is the PU signal waveform, and $n(t)$ is a zero-mean additive white Gaussian noise (AWGN).

Having (12) and (13), considering a network having C channels and M PUs, the bandwidth of each SU can be calculated as follows:

$$R(t) = e^{-\theta_1 \frac{T}{\tau_{on}} \frac{C}{M}} R_s(t) + (1 - e^{-\theta_1 \frac{T}{\tau_{on}} \frac{C}{M}}) R_f(t) \tag{14}$$

where θ_1 is the scaling factor and τ_{on} is the time that PUs are ON state.

Considering (11) and (14), each CH calculates a weight for its own cluster and includes it in the CHRReq. Defining CH_{BW_i} as a bandwidth of cluster i, the cluster routing weight (CRW) is formulated as follows:

$$CRW_i = (\omega).\frac{1}{\mu} + (1 - \omega).CH_{BW_i} \tag{15}$$

where ω is the assigned weights (priority) given to the μ or CH_{BW_i}. From (7) and (15) we can have

$$M(r) = \sum_{i=1}^{N} (\omega).\frac{1}{\mu} + (1 - \omega).CH_{BW_i} \tag{16}$$

References

[1] FCC Spectrum Policy Task Force, "Report of the spectrum efficiency working group," 2002.

[2] G. Staple and K. Werbach, "The end of spectrum scarcity [spectrum allocation and utilization]," *Spectrum, IEEE*, vol. 41, no. 3, pp. 48–52, 2004.

[3] R. Tandra, S. M. Mishra, and A. Sahai, "What is a spectrum hole and what does it take to recognize one?" *Proceedings of the IEEE*, vol. 97, no. 5, pp. 824–848, 2009.

[4] J. Mitola and G. Q. Maguire Jr, "Cognitive radio: making software radios more personal," *Personal Communications, IEEE*, vol. 6, no. 4, pp. 13–18, 1999.

[5] I. F. Akyildiz, W.-Y. Lee, M. C. Vuran, and S. Mohanty, "Next generation/dynamic spectrum access/cognitive radio wireless networks: a survey," *Computer Networks*, vol. 50, no. 13, pp. 2127–2159, 2006.

[6] I. F. Akyildiz, W.-Y. Lee, and K. R. Chowdhury, "Crahns: Cognitive radio ad hoc networks," *Ad Hoc Networks*, vol. 7, no. 5, pp. 810–836, 2009.

[7] Q. Zhao, S. Geirhofer, L. Tong, and B. M. Sadler, "Opportunistic spectrum access via periodic channel sensing," *Signal Processing, IEEE Transactions on*, vol. 56, no. 2, pp. 785–796, 2008.

[8] I. F. Akyildiz, W.-Y. Lee, M. C. Vuran, and S. Mohanty, "A survey on spectrum management in cognitive radio networks," *Communications Magazine, IEEE*, vol. 46, no. 4, pp. 40–48, 2008.

[9] R. Peeters, "The maximum edge biclique problem is np-complete," *Discrete Applied Mathematics*, vol. 131, no. 3, pp. 651–654, 2003.

[10] D. E. Knuth, *The art of computer programming: sorting and searching*. Pearson Education, 1998, vol. 3.

[11] R. T. Marler and J. S. Arora, "The weighted sum method for multi-objective optimization: new insights," *Structural and multidisciplinary optimization*, vol. 41, no. 6, pp. 853–862, 2010.

[12] N. Bulusu, D. Estrin, L. Girod, and J. Heidemann, "Scalable coordination for wireless sensor networks: self-configuring localization systems," in *International Symposium on Communication Theory and Applications (ISCTA 2001), Ambleside, UK*, 2001.

[13] G. Shah, O. B. Akan *et al.*, "Csma-based bandwidth estimation for cognitive radio sensor networks," in *New Technologies, Mobility and Security (NTMS), 2012 5th International Conference on*. IEEE, 2012, pp. 1–5.

Social Adoption of ICT Based Healthcare Delivery Systems in Rural Bangladesh

(A Case Study on Portable Health Clinic)

Md. Nazmul Hossain
Dept. of Advanced Information Technology
Kyushu University, Fukuoka, Japan
nazmul@f.ait.kyushu-u.ac.jp

Kazi Mozaher Hossein
Dept. of Advanced Information Technology
Kyushu University, Fukuoka, Japan
kmhjewel@f.ait.kyushu-u.ac.jp

Rajib Chakrabarty
Global Communication Center
Grameen Communications, Dhaka, Bangladesh
rajib@mail.grameen.com

Hiroshi Okajima
Toyota Motor Corporation, Aichi, Japan
hiroshi_okajima_aa@mail.toyota.co.jp

Hironobu Kitaoka
Toyota Motor Corporation, Aichi, Japan
hironobu_kitaoka_aa@mail.toyota.co.jp

Ashir Ahmed
Dept. of Advanced Information Technology
Kyushu University, Fukuoka, Japan
ashir@ait.kyushu-u.ac.jp

ABSTRACT

ICT based healthcare services: mHealth and eHealth in particular have gained huge momentum around the globe including under developed economies due to the high penetration of mobile phone and easy cum inexpensive access to information and communication technologies even among rural communities. Bangladesh is no exception. Till March 2012, a total of 26 mHealth and/or eHealth initiatives have been taken, many of which are still straggling without having any major technical flaw. Although there is ample evidence and concerns pertaining to the technological perspective of eHealth, there are only a few studies conducted in regards to the social adoption process and consumer behavior of remote healthcare systems. This paper aims to explore and analyze the behavioral aspects of eHealth consumers, upon which social acceptability and business sustainability of ICT based healthcare systems mostly depends. In this research, we have selected two current experimental rural sites of Portable Health Clinic (PHC) namely Bheramara in Kushtia district and Kalihati in Tangail district. The study shows current consumer's trend and health status in terms of age, gender and likely health risk. It also depicts an overall comparison of health status between two of the experimental sites and the movement of health condition with age taking 4969 rural patients as a sample.

Keywords

ICT Based Healthcare Systems; Portable Health Clinic (PHC); Social Adoption; Consumer Behavior; Unreached Community; Rural Bangladesh.

1. BACKGROUND

Bangladesh has a serious shortage of physicians, paramedics, nurses, and midwives. The nurse–physician ratio is one of the poorest in the world. There are approximately three physicians and one nurse per 10,000 people, the ratio of nurse to physician being only 0.4 [1]. The available qualified healthcare providers are centered in urban areas while the majority of people live in rural areas, resulting in an inequitable access to quality healthcare for the rural and disadvantaged sections of the population [2]. Under these circumstances, ICT based healthcare services i.e. mHealth and eHealth have been emerged in Bangladesh since late 90's which provides a new opportunity to ensure access to quality healthcare services for the population in general, and for people from poorer sections and hard-to-reach areas in particular. Effectiveness of these services depends on the evidence-informed development of appropriate programs designed around people's perceptions of ICT based healthcare systems and user feedback.

1.1 ICT Based Healthcare Systems in Bangladesh

Access to quality health services and associated costs are a threat to Bangladesh's current momentum for universal health coverage (UHC). The existing health system is largely (>60%) dependent on out-of-pocket payments [3]. Among many health system concerns, a serious lack and unequal distribution of qualified health human resources (HHR) is a harsh reality. Only 25% of the HHR is working for the rural population which accounts for 70% of the total population [4]. The health system of Bangladesh is haunted by challenges of accessibility and affordability. Despite impressive gains in many health indicators, recent evidence has raised concerns regarding the utilization, quality and equity of healthcare. In the context of new and unfamiliar public health challenges including high population density and rapid urbanization, eHealth and mHealth are being promoted as a route to cost-effective, equitable and quality healthcare in Bangladesh [5].

eHealth and mHealth have been defined in many ways that essentially confer more or less similar attributes. eHealth is an umbrella that includes a spectrum of technologies including computers, telephony and wireless communications to provide access to health care providers, care management and education [6]. mHealth is essentially a subset that delivers such services via mobile phones [7]. In brief, eHealth and mHealth facilitate provision of healthcare through information and communication technology. Globally, eHealth is steadily becoming a popular platform for healthcare delivery and Bangladesh is no exception. A number of initiatives have already been implemented since the late 90's. These have mainly focused on mobile phones, especially important amongst the rural and underserved communities for their potential to overcome geographical boundaries. In 2011, WHO reported Bangladesh as one of the 15 countries using mHealth to raise health awareness [8].

The year 1998 is a milestone for eHealth in Bangladesh as the first eHealth project was launched by Swinfen Charitable, a not-for-profit institute. It involved a collaboration between the Centre for the Rehabilitation of the Paralyzed (CRP) in Bangladesh and the Royal Navy Hospital Haslar, in UK. During the same year, the Ministry of Health and Family Welfare (MoHFW) initiated their first eHealth initiative [9]. Just a year later the Telemedicine Reference Center Limited (TRCL), a private company, initiated the use of mobile phones for healthcare delivery. In 2001, a professional coalition was established, the Bangladesh Telemedicine Association (BTA). This provided a platform for the ongoing and sporadic eHealth initiatives in the country. A similar platform called the Sustainable Development Network Program (SDNP) was formed in 2003, aimed at establishing better collaboration and understanding between providers [4]. Later in 2006, TRCL paired with GrameenPhone (GP) and initiated a mobile phone based call center for subscribers called Health Line: 789 [10]. A number of NGOs, including BRAC, Sajida Foundation and DNet subsequently developed an interest in eHealth and mHealth. The main focus of their interest was on enhancing the efficiency of project implementation, for example by monitoring and evaluating interventions. Later many private entities became involved in telemedicine and/or patient record systems in their clinics and hospitals.

A study conducted by ICDDRB identified, till March 2012, in total 26 initiatives (either pilot or full scale programs) with direct or indirect associations with eHealth and/or mHealth in Bangladesh among which four were public, eighteen private and four NGO [5].

1.2 About Portable Health Clinic (PHC)

The Portable Health Clinic (PHC) system is an e-health system with a telehealth component. The PHC was jointly designed by the 'Social Technology Lab' at Kyushu University, Japan and Grameen Communication's Global Communication Center (GCC) to provide affordable e-Health service to low-income, low literate people living in unreached communities [11].

The PHC back-end comprises GramHealth software applications, database, and medical call center. GramHealth software applications process patients' Electronic Health Records (EHR) and doctor's e-Prescriptions, and store them in a database. Doctors at the medical call center access GramHealth data cloud through the Internet or have a copy of the database in the call center server. Upon receiving a multimedia call from a patient, the doctor can find patient's previous EHR, can create, and send an e-Prescription [12]. This saves time and effort as the doctor does not need to ask questions about the patients' personal profile (basic attributes and medical history) but can focus on the immediate health inquiry.

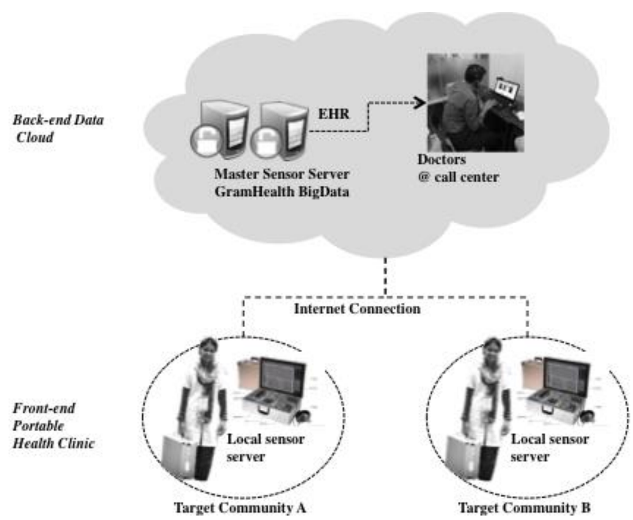


Fig. 1. Portable Health Clinic (PHC) system architecture

The PHC front-end has the instances of portable briefcase consisting of medical sensors and measuring equipment operated by healthcare workers living in unreached communities. The medical sensors are used to identify non-communicable diseases (NCDs) and have a Body Area

Network' (BAN) interface to transmit patient data to tablet PC (local sensor server within the briefcase). The local sensor server synchronizes its cache with the master sensor server when an Internet connection is available. The master sensor server in the back-end data cloud stores all sensor data and provides data to the GramHealth database and doctors in the call center [12].

2. RESEARCH MOTIVATION AND OBJECTIVES

Since the ICT based healthcare services are comparatively new to the low income, low literate rural people of Bangladesh, the sustainability of PHC will largely depend on consumers perception and social acceptance of new technology. Therefore, the understanding of factors that influence technology acceptance is essential for its successful adoption [13]. Although there is ample evidence and concerns pertaining to the technological perspective of eHealth, there are only a few studies conducted in regards to consumer adoption [14]. It is therefore important to measure the attitudes of consumers towards eHealth systems and the system characteristics, which directly affect system acceptance once implemented [15].

The primary objective of the research is to explore and examine the consumers' trend and behavioral factors that affect social adoption of ICT based healthcare services in rural Bangladesh.

The secondary objectives of the study are as follows:

- To assess the current trend and health status of rural patients served by Portable Health Clinic (PHC) in two experimental sites.
- To investigate the current level of knowledge and awareness of rural people on ICT based healthcare systems (PHC in particular).

3. RESEARCH METHODOLOGY

Social adoption of a new product or service, specifically means the large scale acceptation of that product or service by its target consumers. We have identified six basic steps that PHC will have to pass through on the way of its overall consumer acceptance.

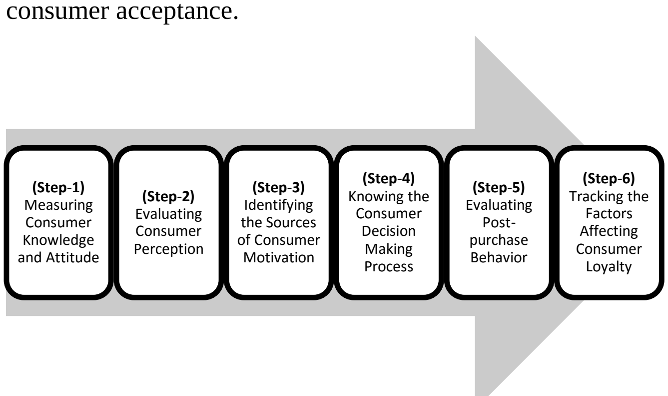


Fig. 2. Steps of PHC's Consumer Acceptance

Now we are in the very first step where the researchers are going to explore and examine the behavioral aspects rural consumers in two target areas regarding their knowledge and attitude towards PHC that affect its social adoption largely.

The study will be qualitative and exploratory in design. Being a qualitative study, data will be generated from personal interviews, field observations and from the examination of documents from various sources. Then the data will be analyzed through appropriate statistical tools. The interviewees will be asked questions (both close ended and open ended) about ICT based healthcare services and the structured questions, eventually generated the stories they experienced in real life. How their lives have been affected and what they experienced by taking the services will be the main theme of the interview. The interview will be audio recorded and later on, transcripts will be written. To maintain the right of privacy of the respondents, they will be briefed on the research purpose and asked whether they want to participate in the study as well use their names and other information in the paper.

In order to achieve the research objectives following information are needed:

Types of Information Needed	Source of Information		
	PHC Database	Questionnaire Survey	Further experiment
1. Overall service status of PHC	✓		
2. PHC consumer distribution by: gender & age	✓		
3. Patient's health status by: gender, age & location	✓		
4. Movement of community health status	✓		
5. Consumer's knowledge about the use of ICT in healthcare		✓	
6. Existing awareness about PHC among target consumers		✓	
7. Use of PHC in recent sickness		✓	
8. Compliance with prescription		✓	
9. Cost perception on PHC and traditional healthcare		✓	
10. Reasons for using and not using PHC		✓	
11. Intention to use PHC in future			✓
12. Measuring the changes in consumer attitude after follow-up call from doctor			✓

After collecting information on the behavioral aspects of PHC consumers, a dedicated Technology Acceptance Model (TAM) will be designed for PHC in order to identify the exact external environmental factors and consumer's behavioral factors that affect long run social acceptance of this new ICT based mobile healthcare service – i.e. PHC.

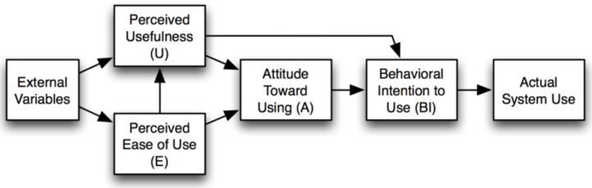


Fig. 3. Technology Acceptance Model (TAM)

Source: Davis, F. D.; Bagozzi, R. P.; Warshaw, P. R. (1989), "User acceptance of computer technology: A comparison of two theoretical models", Management Science 35: 982–1003

4. SURVEY PROFILE

Portable Health Clinic (PHC) has started serving actively since 2012, till March 15, 2016 it reached 32 remote locations in 9 districts and served 33,283 rural patients among which 18,372 (55.2%) were male and 14,911 (44.8%) were female.

For this research, we have selected two communities (sub-districts): Bheramara sub-district in Kushtia District and Kalihati sub-district in Tangail district. Bheramara consists of 6 different unions where the total number of households is 15,698 and total population is 176,618. Males constitute 51.55% of the population, and females 48.45%. Bheramara has an average literacy rate of 28.1% (7+ years), and the national average of 32.4% literate. Kalihati consists of 13 different unions where the total number of households is 50,982 and total population is 398,786. Males constitute 51.55% of the population, and females 48.45%. Kalihati has an average literacy rate of 42.9% (7+ years).Male literary rate is 46.1% and female literary rate is 39.8 %

In this two experimental sites we have a total of 66,680 households among which we are going to randomly select 5,628 respondents (not more than one from a household). In selecting this sample size we considered 95% confidence level and the confidence interval was 1.25

5. CONSUMER TREND AND HEALTH STATUS

PHC has a database of 33,283 rural patients who have been served so far among which a total of 5012 patients are from our two experimental sites namely Bheramara and Kalihati. PHC started serving in those two experimental sites from 2013. Among this total 5012 patients 3041 (60.7%) were male and 1971 (39.3%) were female.

5.1 Overall Performance

		Bheramara		Kalihati	
Year	Active Month	Yearly Service	Service Per Month	Yearly Service	Service Per Month
2013	5	801	160	517	103
2014	9	837	93	483	54
2015	12	1130	94	975	81
2016	2.5	483	193	376	150
Total		**3251**		**2351**	

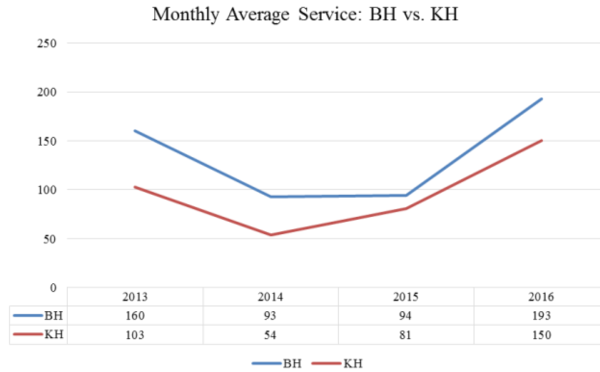


Fig. 4. Service Performance: Bheramara Vs. Kalihati

The above figure shows that, PHC is serving more patients in Bheramara than Kalihati. Kalihati has a location advantage which is close to the capital city Dhaka and another big city Gazipur this is why the people of Kalihati have comparatively better access to traditional healthcare services. On the other hand Bheramara is a remote village having less access to healthcare services which made PHC more acceptable in this area.

5.2 PHC Consumers by Age

Age	Frequency	Percentage
0-19 (Childhood)	123	2.5%
20-39 (Young Adult)	1874	37.5%
40-64 (Middle-aged Adult)	2485	49.8%
65-84 (Senior Age)	489	9.8%
85+ (Old Senior Age)	23	0.5%
Total	4994	100%

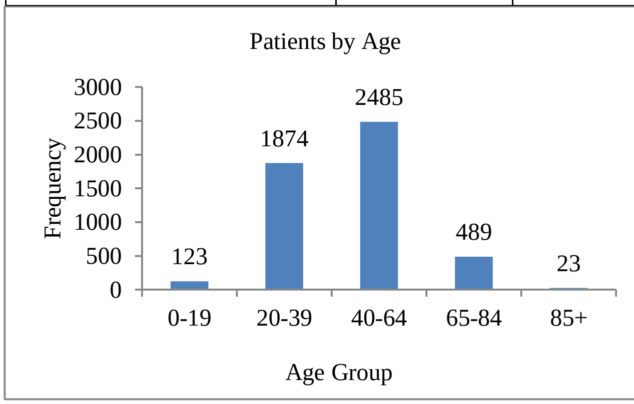


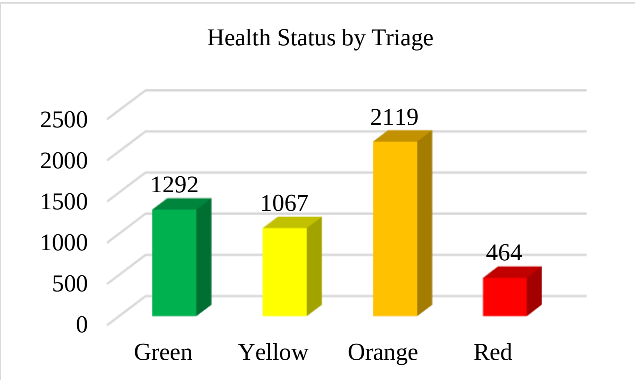
Fig. 5. Age Distribution of Patients

The biggest consumer segment of PHC is Middle-aged Adult (40-46) followed by Young Adult (20-39)

5.3 Overall Health Status

On the basis of some measurable health condition like blood pressure, pulse rate, blood glucose, arrhythmia etc. PHC has classified its patients into four different color groups where green indicates safe, yellow for doubtful, orange means going to be in danger and red indicates the patients who are already in danger zone.

	Green	Yellow	Orange	Red	Total
Male	819	652	1259	267	2997
Female	473	415	860	197	1945
	1292	1067	2119	464	4942



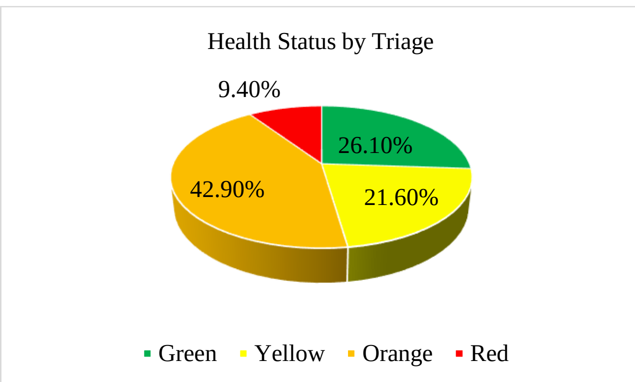
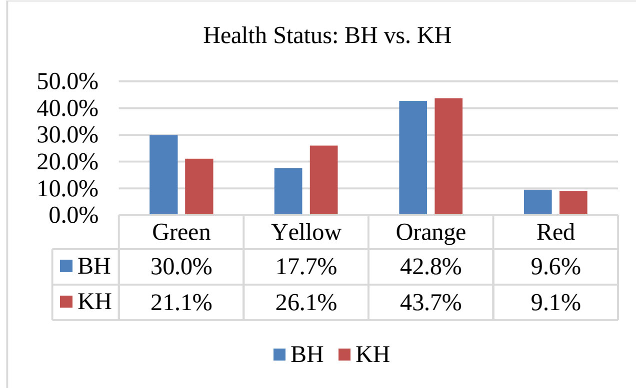


Fig. 6. Health Status of Experimental Sites

5.4 Comparative Health Status: BH vs. KH

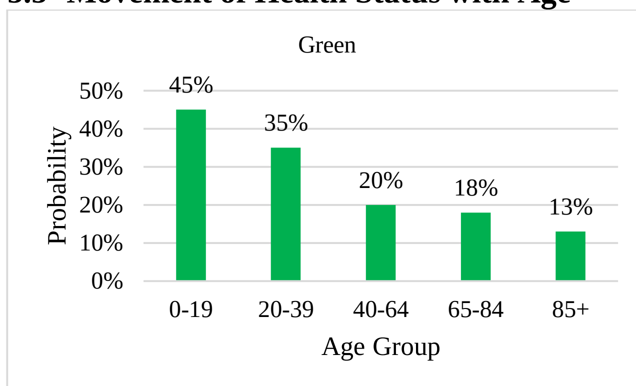
	Green	Yellow	Orange	Red	Total
BH	821	486	1172	262	2741
KH	471	581	974	202	2228

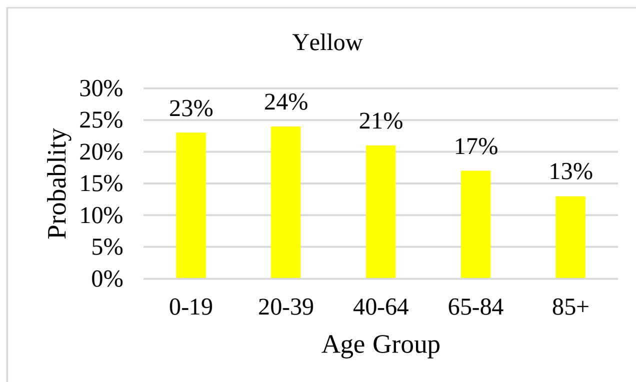
4969

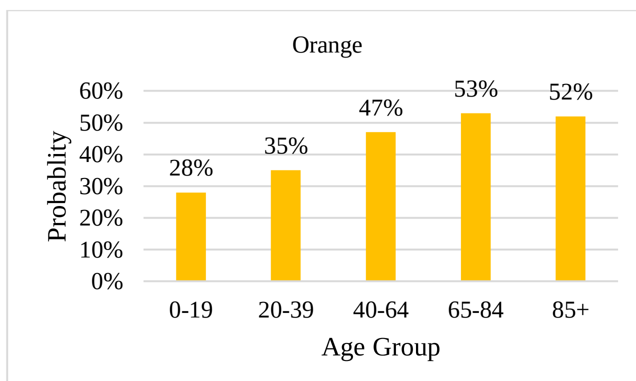


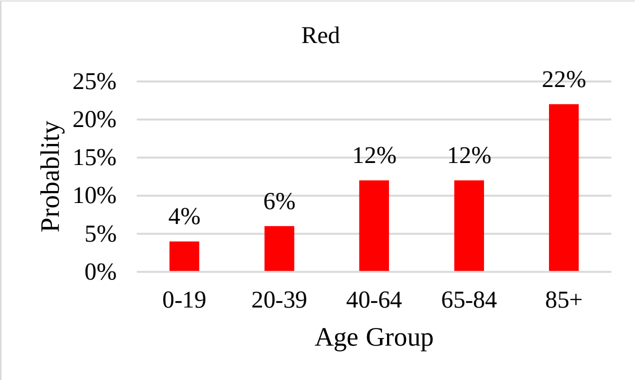
“Green” is significantly higher in Bheramara while “Yellow” is higher in Kalihati.

5.5 Movement of Health Status with Age





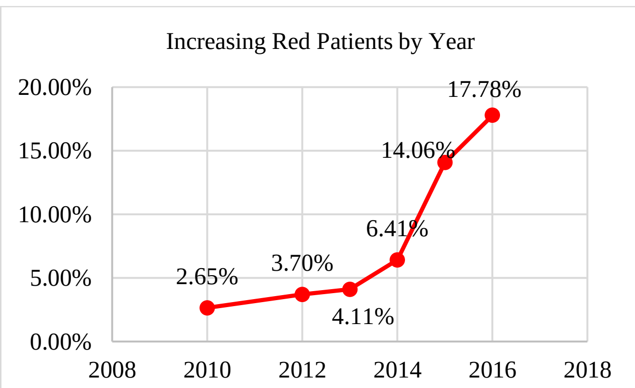




The above graphs clearly shows as age goes up probability of being “Green” and “Yellow” goes down while the probability of being “Orange” and “Red” goes up with age.

5.6 Social Adoption of PHC

Year	Patients Served	Red Patients	% of Red Patients
2010	377	10	2.65%
2012	6412	237	3.70%
2013	13293	546	4.11%
2014	5384	345	6.41%
2015	5713	803	14.06%
2016	2104	374	17.78%



The above figure shows PHC is serving more vulnerable patients every year which indicates more reliability and social acceptance of PHC in its service area year by year.

6. WHAT NEXT?

A structured questionnaire survey will be conducted among the target sample in two experimental sites in order to investigate the current level of knowledge and awareness and other behavioral factors that affect social adoption of Portable Health Clinic.

Primary data collected from the survey will be analyzed through ANOVA, Correlation and Regression analysis in order to find out the pattern and intensity of relationships between the dependent variable i.e. social adoption of PHC and other independent variables i.e. demographic, behavioral and social traits of the consumers.

7. EXPECTED OUTCOMES

After analyzing the survey results the researchers will be able to identify the rural consumers’ knowledge and understanding about the use of ICT in healthcare, the existing level and sources awareness towards PHC among its target consumers, use of PHC in recent sickness, compliance with prescription suggested by PHC call center doctors, cost perception on PHC and traditional healthcare and the reasons for using and not using PHC.

8. CONCLUSION

The high penetration of mobile phone and easy access to information and communication technologies into the society, even among rural communities, provides a unique opportunity to use the eHealth technology for taking healthcare to the people in general and the ones with difficulty in accessing the services involving a physical visit to a physician in particular. It is a prerequisite that knowledge be universal to ensure exploitation of the full potential of the services. So, this is high time to explore the consumer behavior, their knowledge and attitude towards ICT based healthcare services. So far we have found the consumer’s trend and health status in terms of age, gender and their likely health risk. We have also shown the overall comparison of health status between two of the experimental sites and the movement of health condition with age. In the next level we will be able to find consumer’s demographic and behavioral factors that affect social adoption of PHC. If the research can be conducted properly, the findings will help ICT based healthcare service providers to design and develop their services in accordance with the perceptions and expectations of their target consumers, i.e. patients. This will also enhance the possibility of wide-spread social adoption and sustainable operation of ICT based healthcare services.

9. ACKNOWLEDGMENTS

The researchers would like to thank Social Technology Lab members of Kyushu University for providing their valuable comments during research discussions; Grameen Communications for gathering and sharing experimental data from field and Toyota Motor Corporation, Japan for funding this research project.

10. REFERENCES

[1] Ahmed, S. M., Hossain, M. A., Chowdhury, A. M. and Bhuiya, A. U. 2011. *The health workforce crisis in Bangladesh: shortage, inappropriate skill-mix and inequitable distribution.* Human Resources for Health 2011 9:3. http://www.human-resources-health.com/content/9/1/3

[2] Iqbal, M. 2013. *Knowledge of, Attitude towards, and Use of mHealth Services in Chakaria, Bangladesh.* Future Health Systems. The UK Department for International Development. December, 2013.

[3] Report BHW: MovingTowards Universal Health Coverage. Dhaka: James P Grant School of Public Health, BRAC University; 2011.

[4] Nessa, A., Ameen, M.A., Ullah, S., and Kwak, K. *Applicability of telemedicine in Bangladesh current status and future prospects.* 2010. The International Arab Journal of Information Technology 2010, 7:2.

[5] Ahmed, T., Lucas, H., Khan, A. S., Islam, R., Bhuiya, A. and Iqbal, M. *eHealth and mHealth initiatives in Bangladesh: A scoping study.* 2014. BMC Health Services Research 2014, 14:260. http://www.biomedcentral.com/1472-6963/14/260

[6] DeLuca, J. M., and Enmark, R. *E-health: the changing model of healthcare.* Front Health Serv Manag 2000, 17(1):3–15.

[7] Mechael, P.N. *The case for mHealth in developing countries.* Innovations: Technology, Governance, Globalization 2009, 4(1):153–168.

[8] Kay, M., Santos, J., and Takane, M. *mHealth New horizons for health through mobile technologies.* In Global Observatory for eHealth series, Volume 3. Geneva: World Health Organization; 2011.

[9] Boutilier, M. *A Survey of E-Health Initiatives Across the Commonwealth.* Geneva, Switzerland: Commonwealth Health Ministers Meeting; 2008. http:// www.secretariat.thecommonwealth.org/files/178273/FileName/ HMM_G__08_6-EHealth.

[10] Telemedicine reference center LTD. achievements in 12 Years (1999–2011). http://www.amcare24.com/download/TRCLCompanyProfile_9APR2011.pdf.

[11] Ahmed, A., Inoue, S., Kai, E., Nakashima, N. and Nohara, Y. *Portable Health Clinic: A pervasive way to serve the unreached community for preventive healthcare.* Proceedings of the 15th International Conference on Human-Computer Interaction (HCI 2013)

[12] Kato, S. *A Study on Implementing a Portable Clinic based on Social Needs.* Undergraduate Thesis, Kyushu University, March, 2012

[13] Randike, G., Renato, I. and Tony, S. 2014. *Consumer Acceptance of AccountableeHealth Systems.* e-Health – For Continuity of Care. European Federation for Medical Informatics and IOS Press. DOI= 10.3233/978-1-61499-432-9-980

[14] Petkovic M, Ibraimi L. Privacy and Security in e-Health Applications. 2011. In: E-Health, Assistive Technologies and Applications for Assisted Living: Challenges and Solutions [Internet]. [23-48].

[15] Or CKL, Karsh BT. A systematic review of patient acceptance of consumer health information technology. J Am Med Inform Assoc. 2009; 16(4):550-60.

Improve Payment Card Security by Adding Voice and Fingerprint Biometric Solution

Md. Tahajul Islam
CSE Department
University of Asia Pacific
Dhaka, Bangladesh
tahajul.islam@gmail.com

Mehedi Mahmud
CSE Department
University of Asia Pacific
Dhaka, Bangladesh
mehedimahmudcse@yahoo.com

Muhammad F. Mridha
CSE Department
University of Asia Pacific
Dhaka, Bangladesh
firoz@uap-bd.edu

***Abstract*—Nowadays Payment Card have been extensible used for remote or onsite transactions to reduce the need for awkward of most cash transaction in human life and people gradually dependent to use payment card like a debit card, a credit card or an ATM card, which are easy to carry, handle and manageable and allows them to immediate fund transfer. This is why Payment cards become the target to the frauds and have been a large number of incidents of payment card frauds due to the security weakness of this system. At present security of the most payment card is guaranteed by PIN (Personal Identification Number), which is shareable with others, on other hand fingerprint and voice biometric of every person is unique and cannot be shareable. In this paper, the author proposed a solution to improve the better performance in payment card security by enhancing the fingerprint and voice biometric.**

Keywords—Fingerprint, voice biometric recognition, PIN, credit card, debit cards, ATM cards.

I. INTRODUCTION

Financial organizations use different security features in payment card system to protect the frauds. In order to assign such separate security measures these organizations invest the huge amount of money and human resources, but because of increasing the usages of organized and sophisticated techniques by the Fraudsters, fraud is still increasing. Therefore, financial organizations, as well as the cardholder, are very worried about the security measure of payment card system. This paper analyzes the necessity of using fingerprint and voice biometric recognition to integrate with payment card system as a replacement of general memorized PIN (Personal Identification Number) verification. There are a lot of factors that could be used for human identification, which is unique for each person. The fingerprint was the first one to be discovered and analyzed. Finger carries a unique pattern which consists of different spirals, loops and curves and which is totally unique. Other factors, which are unique for each person, include a face, hand and palm geometry, eye iris and retina, ear shape, handwriting and DNA. Some techniques using these factors are still under development which may require new technologies and devices to detect the deviations. For communication purpose, many technologies are developed for speech generation, speech recognition, special devices and character recognition such as OCR applications which make an easier operation of devices. Speech recognition is an easy technology which needs only a single microphone to get the necessary sample where people provide their speech sample to the system very willingness. The biometric factors could be divided into physical (such as fingerprint, hand, iris, retina, face, and palm) and behavioral (such as voice, movement characteristic and signature). Here, the Hidden Marcov Model (HMM) and different search algorithm along with lexical analysis are mediated [1] so that the significance in payment card security rise to a completely new level. In this report, the research initially discusses study rationale and latest studies in this context. The researcher also discusses the anti-fraudulent activities. The analysis and experiments are prescribed in the later segment and based on these, the conclusion is drawn.

II. MOTIVATION AND STUDY RATIONALE

Recently, the frauds in payment cards increase such a rationally which have threatened the management of information security. Fraudsters are using some new well organized and sophisticated techniques which are the biggest challenge to ensure the security level of payment system for every user of all environments. To improve this security, the integration of both fingerprint and voice biometric speech recognition technique can be used as an unparalleled and high-level enhancement in the aspect of payment card frauds. Though both fingerprint and voice biometric are treated as a unique identification for each person, so fraudster must be hacked this two high-level identification which will be very difficult in the aspect of committed the payment card transactions.

III. LITERATURE REVIEW AND METHODOLOGY

A. Literature Review

Biometric technique is becoming the most popular method in highly secure identification and verification. Though the level of security are breach and transaction incident are increase, the necessity of highly secure identification and verification technologies are become ostensible. There are various types of biometrics in this world such as fingerprint, hand geometry, retinal scanning, signature verification, iris scanning, voice recognition and facial recognition.

B. Fingerprint Recognition

Though every person has a unique characteristics and patterns, so in biometric technology, fingerprint is the most usages technology to identify or verify of any individual person. A fingerprint pattern is consist of various space and lines, where lines are refereed as ridges and the spaces between these ridges are called valleys, both valleys and ridges are needed to match for identification and verification. Any comparisons of the fingerprint are made on the unique fingerprint traits which are called **minutiae**. The general live scan produces 40 minutiae [2].

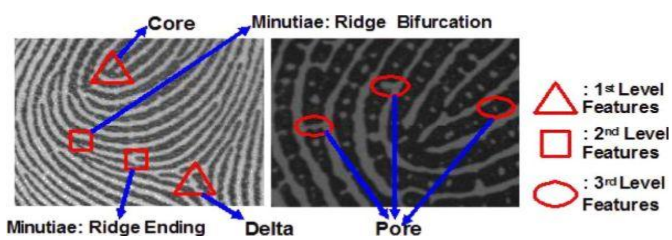


Fig 1: Three level feature of Fingerprint.

According to the principal of fingerprint recognition, all fingerprints could be divided into 5 classes which are Arch, Left Loop, Right Loop, Tended Arch and Whorl.

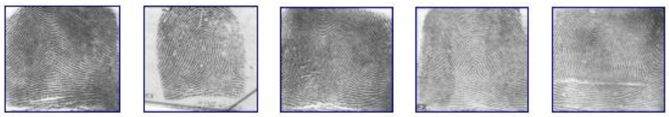


Fig. 2. Fingerprint classes- Arch, Left Loop, Right Loop, Tended Arch and Whorl.

There are 5 steps in process of minutiae comparison (fingerprint scan analysis, identification and verification) consists of Acquisition of fingerprint image, Image Processing, Location of Unique characteristics, Template generation and Template matching. In the stage Acquisition of fingerprint image its mainly involves the image processing, such as scaling, which is a process of improving the appearance of the acquired image and the results in a series of thick black ridges contrasted to white valleys[3]. In this process image features are recognized and enhanced for verifying against the pre stored minutiae file. Enhancement of the acquired image is used to minimize the characteristics of fingerprint ridges and valleys from unique patterns such as Arch, Loop and Whorl. Template is generated after the identification of the unique characteristic of finger by Template generation step and finally identification is done by the Template matching step.

Finger Sample → Acquisition of fingerprint image → Image Processing → Location of Unique characteristics → Template generation → Template matching

Fig. 3. Process of minutiae comparison

C. Speech Recognition

Voice biometric recognition is the capability of a program or machine to understand or receive and convey the spoken instructions, which is one of the safe and convenient recognition techniques [5]. This technique combines both physical and behavioral factors to generate speech patterns which can be captured by the voice processing technology. Speaker's inherent properties such as fundamental frequency, nasal tone, inflection, cadence etc. are used for voice recognition systems, which is dependent on the generated waveforms and air pressure patterns while a user speak. The microphone captures the speakers' voice signal and converts this analog signal into digital signal. Due to unsatisfactory accuracy voice recognition technique is still unreliable and some throat related problems speakers' voice may be rejected by this system.

Here the author uses the Hidden Marcov Model which is a popular algorithm to implement speech recognition [6]. Time variants under discrete states are identified. The min_HMM (minimal Hidden Marcov Model) in C language recognize speech. This model has two portions, one the front end and other is the search algorithm. In the front end, there contains an analog to a digital encoder. Afterward, the digital speech is sampled and analyzed under spectral shaping analysis. The message is then transformed using the finite impulse response (FIR) filter. The Signal processing is as follows:

Message → Analog to Digital Encoder → Spectral Shaping → Spectral Analysis → Parametric Transformation → Search Algorithm

Fig. 4. Search Algorithm

The search algorithm uses an observation sequence and analyzes with the preserved word and the mathematical model in this regard as follows:

$$\max p(\omega_i|O),$$

where ω_i is the probability of word and O is the observation vector.

To recognize the voice of any speaker the following steps may be followed:

Voice Capturing: This is the first steps where speakers' voices are captured by a quality full Microphone with a sampling frequency at least of 11 KHz and a precision of 16 bits and converts this analog signal into digital signal. The result may be differing according to the capturing quality of this recognition.

Signal Pre-processing : The next steps is the pre-processing of the captured signal which consists of various elementary steps such as pre-emphasis, framing, windowing and clipping of the non-speech frames, i.e. selecting of the speech frames. Feature extraction: The 3rd step is the feature extraction from the processing signal, where the actual frames have been extracted from the signal and are ready to the next processing. The goal of this stage is to extract some typical features from the signal. There are a lot of various possibilities and it is very difficult to select the best of them. The features are based upon some parameters of the captured signal. These parameters may be e.g.: Zero Crossing Rate (ZCR), autocorrelation, Linear Prediction Coefficients (LPCs), energy, Mel-Frequency Spectral Coefficients (MFSCs), Coefficients (MFCCs), Mel-Frequency Cesptral see [7], [8]. These parameters are not sufficed for the voice recognition, so it is necessary to extract form the captured signal some other information which enable us to recognize the voice reliably. Recognition: This is the final step of speakers' recognition. The result (confirmation of success or refusal) depends before all on the feature set chosen for this method and then, even the worst recognition method would be able to execute the recognition properly. To recognize the speaker and extract a unique feature vector we may need to user some tolls like the hidden Markov models or the neural networks.

IV. ANALYSIS AND WORK FLOW

The idea of this proposed system is to secure the transaction of any payment card system which has become a significant fact in case of security. In this paper the author analyzed the necessity of using fingerprint and voice biometric recognition to integrate with the payment card system as a replacement of general memorized PIN (Personal Identification Number) verification.

The user shall use the fingerprint sample and voice sample (passcode) instead of PIN here. The fingerprint and voice sample are preserved when registration takes place. First the user provides the fingerprint and voice sample which are converted digitally. Then the digital message is sampled and matched using search algorithm. In this regard, the min HMM is used.

A. Work Flow Diagram

The proposed workflow diagram is shown in Fig. 6. In this system, no extra microcontroller or accuracy is stored. The matching is conducted in server side. Another java based simulator is installed which send data as soon as the card is inserted in the POS or ATM machine. To make a financial transaction cardholder must Swipe the card first. The system will verify the card and after successful verification, the system will prompt for cardholder fingerprint when cardholder put his/her fingerprint system will match this to the stored sample against the given card number. After successful authentication of fingerprint system will then prompt for cardholder voice biometric (voice passcode), after capturing the voice passcode system will match this to the stored sample against the given card (if the card is EMV them system can authenticate the both fingerprint and voice passcode from the memory of card chip, but for magnetic card system will authenticate those from the server side). After successfully authentication of voice passcode transaction will be committed, otherwise, transaction will be failed.

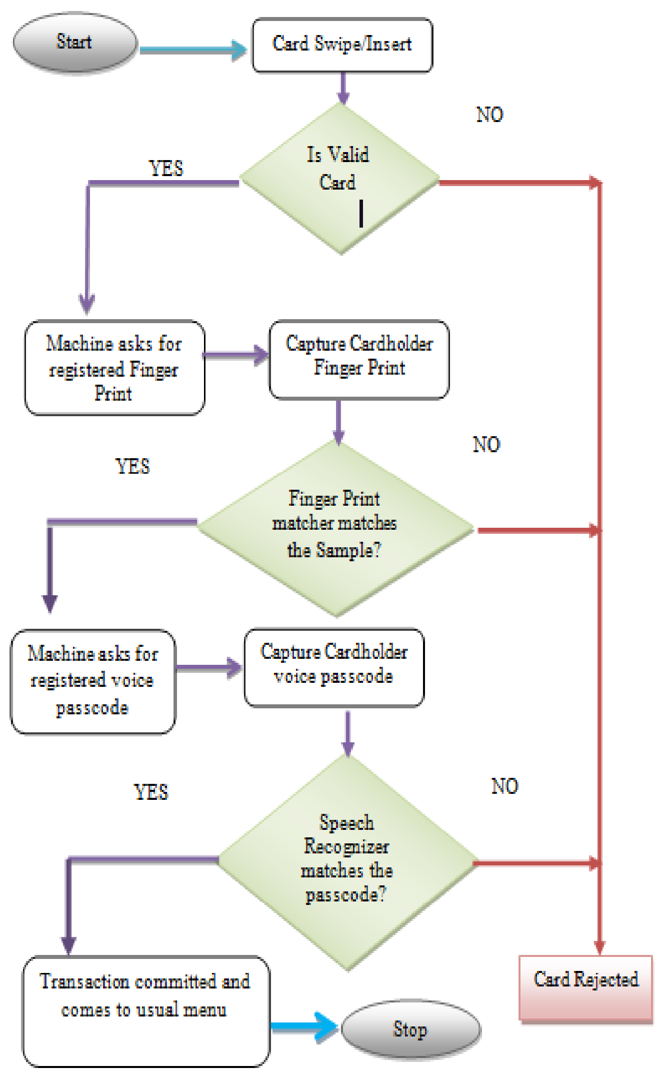


Fig. 5: Flowchart of proposed System

V. EXPERIMENTALEVALUATION

In this section, the author evaluates the proposed security system. In order to testify the system, the user takes 7 subjects and a computer where the usual Card Reader simulator is installed. The purpose of using a computer is that the Card Reader can be simulated in computer and in this simulator virtual card can be swiped. The computer is connected to a local LAN network and is able to talk with oracle 11g database where the subjects' registered Finger Print and voice passcode is stored against the cardholder. A microphone is used gathers voice passcode and a fingerprint capture machine is also installing into the computer. To convert voice passcode from analog to digital sample the ADCPro software is used. This sample is tried to match under LAN with the stored passcode. If matches the system gives to do for father operation. In this experiment, four users passed the system, one did not pass due to heavy noise, one used recorded passcode, one used the wrong passcode and one used wrong fingerprint. Hence, the system can be tuned in the case of noise aspects. The comparison table is given below:

Table 1.Result Table

Cardholder	Finger Print (Match Case 75%)	Voice recognition (Match case 70%)	EMV/Magnetic Card?
1	81%	73%	EMV
2	55%	NA	EMV
3	86%	88%	EMV
4	77%	75%	Magnetic
5	6%	NA	EMV
6	0%	NA	Magnetic
7	78%	0%	Magnetic

VI. CONCLUSION

Payment cards system is widely used all over the world via different means. Many people use the card to purchase products in shops and online market. The author has presented the importance of security while using credit cards. The introduction of voice recognition and finger biometric recognition is optimized in this report. The idea of inclusion these techniques are very rigid, so that the card theft is minimized. The author also suggests is to replace the magnetic card with EMV card which not only lessen card theft but also lower costs. The main purpose and benefit of a voice and finger recognition system are to uplift security and avoid unwanted and illegal transactions.

REFERENCES

[1] L. H. Swee and Texux Instrument, "Implementing Speech Recognition Algorithms on TMS320C2xx Platform", Digital Signal processing solution, 1998.

[2] Mohammad Lawman Ahmed "Use Of Biometrics to tackle ATM fraud," International Conference on Business and Economics Research, IACSIT Press, Kuala Lumpur, Malaysia, Vol. 1, pp.: 331-335, 2011.

[3] Arun Ross, Jidnya Shah and Anil K. Jain, "Towards Reconstructing Fingerprints from Minutiae Points", Vol. 5799, pp 68-80, USA, 2005.

[4] Jian Feng, Anil K. Jain, Fellow, "Fingerprint Reconstruction: From Minutiae to Phase", IEEE, Vol. 33, No. 2, pp. 209-223, February 2011.

[5] Baum, L.E., and Petrie, T., "Statistical Inference for Probabilistic Functions of Finite-State Markov Chains," Annotated Mathematical Statistics, Volume 37, pp. 1554−1563, 1966.

[6] J. G˚arding and T. Lindeberg. Direct computation of shape cues using scale-adapted spatial derivative operators. International Journal of Computer Vision, 17(2):163–191, 1996.

[7] Huang, X., Acero, A., Hon, HW.: Spoken Language Processing, New Jersey, USA, Prentice Hall, 2001, ISBN 0-13-022616-5.

[8] Gold, B., Morgan, N.: Speech and Audio Signal Processing, New York, USA, John Wiley & sons, inc., 2000, ISBN 0-471-35154-7.

Artificial Bee Colony based Optimal PMU Placement in Power System State Estimation

Mohammad Shoaib Shahriar[1]*, Farhan Ammar Ahmad[1], Ibrahim Omar Habiballah[1], Mohammed Afzal Asif[2], Shagorika Mukherjee[2]

[1]Department of Electrical Engineering, King Fahd University of Petroleum & Minerals (KFUPM), Dhahran, Saudi Arabia

[2]Department of Electrical and Electronic Engineering, Independent University, Bangladesh

*shoebeee05@gmail.com

***Abstract*— Advanced power system nowadays demands to ensure accurate estimate of the system states (bus voltages and angles) in order to ensure protection, monitoring, controlling and smooth running of the system. The conventional state estimator estimates the power system states based on the measurements obtained from the Supervisory Control and Data Acquisition (SCADA) system. Besides, the use of synchronized phasor measurement unit (PMU) has become a popular choice of the time in this field because of its ability of providing the real time phasors of voltage and currents. For a better phase estimation accuracy of the system, proper site selection of placing PMU is a must. Keeping an eye on system observability and system meter economy, it is important to identify the optimal location and number of PMUs to be used in state estimators. This paper presents an optimal solution of PMU's placement by using one of the most widely used intelligent technique, Artificial Bee Colony (ABC). State estimation is then carried out by placing PMUs on the resultant optimum locations. Both the voltage and current phasors has been taken from the PMUs installed in the buses. The impact of bad data in the measurement series is also investigated. IEEE 14-bus and 30-bus systems has been taken as test system and Weighted Least Square (WLS) Algorithm has been considered as estimator algorithm to carry out the estimation process.**

Keywords*— *State Estimation, Phasor Measurement Unit, Optimal PMU placement, Artificial Bee Colony.

I. INTRODUCTION

State estimation is such kind of an imperative process which does the very important duty of ensuring power system security by monitoring it precisely. A state estimator estimates the values of state variables (voltage magnitudes & phase angles) at the buses after it is provided with the measurements like magnitudes voltages and power injections in the buses, power flows and current flows in the branches [1-2]. Conventional Supervisory Control and Data Acquisition (SCADA) of remote terminal units (RTU) installed at the substations provides the estimator with these measurement values [3]. After introducing Phasor Measurement Unit (PMU) around the year 1990, it has become the most reliable tool in wide area monitory system (WAMS) which provides real time bus voltage and branch current phasor measurements. The very high refreshing rate of PMU has made it highly acceptable in power system control compared with conventional SCADA systems [4]. Besides of state estimation, PMU has got some more applications like transient stability analysis of power system, fault detection, wide area protection, transmission line thermal monitoring etc. [5-7].

For proper running of any estimator, it must provide sufficient amount of measurements that make the system observable. The minimum number of measurements provided must be (2N-1) if N indicates the number of buses. Lots of work has already been done on this field and researchers has used both the heuristic and mathematical algorithms as solution methodologies [8-9]. Integer programming and exhaustive search technique was used to search for best placement location of PMUs which has been explained in detail in the taxonomy work on PMU placement done by Nikolas et al. [10]. Among the heuristic techniques, Tabu search, Simulated Annealing, Genetic Algorithm, Particle Swarm Optimization, Differential Algorithm, Bee colony, Spanning Tree Search, Immune Algorithm, Recursive Security N algorithm and several hybrids of such intelligent techniques were used in Optimum PMU placement (OPP). The recent most survey on OPP by Nazari-Heris et al. [11] presented all such techniques along with the solution sets for each as well as the objective functions considered in those problem formulation. All of those works mainly focused on minimizing the number of PMUs to be installed in the system which will make the system observable as well as will fulfill several other specified objectives.

This paper presents a technique to find out the optimal PMU locations in a power system. As it is known, that a bus if installed with a PMU, it can provide the voltage phasors of that bus as well as the current phasors of all or some of the adjacent branches connected to that bus depending upon the communication facility availability in that bus location. A MATLAB based optimization tool along with Artificial Bee Colony (ABC) is used to solve the optimization problem. Different solution sets are found for OPP which has been further investigated by doing state estimation. Performance of the estimator is checked by placing PMUs in the optimal solution sets keeping the measurement redundancy level same for all which will enable us to compare between the solution sets. Test cases with and without presence of bad data are considered both for estimation. Widely used WLS algorithm is used to carry out the state estimation for IEEE 14-bus system.

In this paper, section II will present the details of optimization problem of PMU placement. Section III will present the two optimization tools used in getting optimal solutions. WLS algorithm will be briefly presented in section

IV. Part V will present all the simulation results with necessary analysis and followed by the conclusion in section VI.

II. PROBLEM FORMULATION FOR OPTIMAL PMU LOCATION

A PMU is different from the conventional meters in the sense that it has got the ability to provide not only the voltage phasors of the connected bus but also the current phasors of the branches which are attached with the bus [12-13]. In this problem formulation, it has been considered that the buses are all equipped with proper communication facilities which will allow an installed PMU on a particular bus to make all its neighboring buses observable. So, the problem formulation of OPP has got the objective to get the minimal set of PMUs so that each of the buses are at least reached once by the installed PMUs.

The objective function of the optimization problem for an n-bus system is [12]:

$$\min \sum_{i=1}^{n} x_i c_i \qquad (1)$$

Where x_i is the decision variable which denotes that whether a bus is installed with a PMU or not. For a 14 bus system, we will have 14 values of x starting from x_1 upto x_{14}. So, x_i can be represented as:

$$x_i = \begin{cases} 1, & if\ PMU\ is\ installed \\ 0, & Otherwise \end{cases} \qquad (2)$$

C_i is the cost needed to install a PMU in a particular bus. It is considered that the installation costs will be same for each PMUs wherever it is installed.

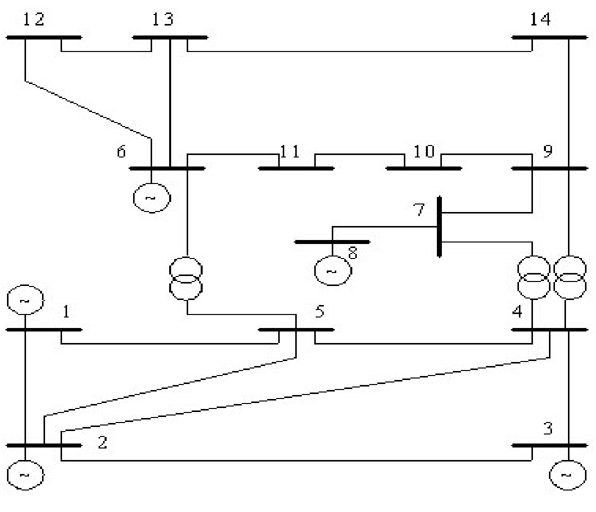


Fig 1: IEEE 14 bus test system

Min $(x_1 + x_2 + x_3 + ... + x_n)$

Constraints of this optimization problem will be used to make the network observable which can be expressed as:

$$A_{nXn}\, X_{nX1} \geq b_{nX1} \qquad (3)$$

Where A_{nXn} is denoted as the connectivity matrix which defines the connection pattern between the buses.

$$A_{i,j} = \begin{cases} 1, & if\ i = j \\ 1, & if\ i\ and\ j\ are\ connected \\ o, & otherwise \end{cases}, \qquad (4)$$

$X_{nX1} = [x_1\ x_2\ x_3\ ...x_n]^T$; a vector with decision variables

And $b = \begin{bmatrix}1\ 1\ 1....1\end{bmatrix}^T_{nx1}$ (5)

For IEEE 14 bus system, the formulation is presented in figure 4.

Subject to $[A]_{14X14}[x]_{14X1} = [b]_{14X1} = \begin{bmatrix}1\\ .\\ .\\ 1\end{bmatrix}_{14X1}$

The connectivity matrix can be seen from the figure and can be formulated from equation (2) as follows:

$$A = \begin{bmatrix}
1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 1 & 1 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0\\
1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0\\
0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0\\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1
\end{bmatrix}$$

From the constraint equation (3), all of the 14 equations can be formulated which will help the system to make the system observable.

Suppose for bus 1 observability equation will be:

$x_1 + x_2 + x_5 \geq 1$; which means that at least one PMU must have to be placed in any of these three buses (1,2,5) so that other two buses could reach the installed PMU with proper communication channel and make the zone around bus 1 observable. Like the same way all other equations are formed for each of the buses which will make the respective bus “reachable” from the neighboring buses. Thus the entire system will become observable by following all the 14 constraint equations.

III. OPTIMIZATION TOOLS FOR OPP

a) CVX:

To solve the convex optimization problems, CVX is widely used which works in Matlab environment. It transforms Matlab into a modeling structured language, allow to incorporate objective functions and constraints using standard Matlab expression syntax which allows the Matlab users to deal with the tool very easily[14].

b) Artificial Bee Colony (ABC)

Artificial Bee Colony (ABC) is a meta-heuristic global optimization algorithm based upon the smart hunting and practical behavior of drone flock. It is a promising method for engineering applications and since its invention in 2005 by D. Karabog, it is being used for numerical optimization, combinatorial optimization problems as well as unconstrained and constrained optimization problems [15]. User defined size of population and cycle number are used as control parameters. Abandonment criteria is also predetermined by the user. The ABC algorithm progresses in mainly four steps:

Initialization

A fixed set of solutions or food source positions are initialized by the scout bees which is denoted as x_m, where the value of m is bounded within 1 to SN (population size). In this step, the controlling limits are also set. Arbitrary food sources produced initially within the predefined range can be acquired from the following equation [16]:

$$x_{mi} = l_i + rand(0,1) * (u_i - l_i) \tag{6}$$

u_i and l_i are the lower and upper bound of x_{mi}. After initialization, the set of solutions are forced to undergo through an iterative search process of the following three phases [17].

Employed Bees Phase

For each employed bee:

- Determine new food source position, v_m using the expression:

  $$v_{mi} = x_{mi} + \phi_{mi}(x_{mi} - x_{ki}) \tag{7}$$

  where x_k is a food source and *i* is a parameter index. ϕ_{mi} is a random number with the range [-a, a].
- Calculate the value fit_m using:

  $$fit_m = \begin{cases} \dfrac{1}{1+f_m} & f_m \geq 0 \\ 1+abs(f_m) & f_m \prec 0 \end{cases} \tag{8}$$

  The m-th fitness value is the fit_m and f_m is the objective function value of the m-th solution.
- v_m's with better fitness values than the corresponding x_m's are used in updating the x_m's. It is done by using a greedy selection.

Onlooker Bees Phase

For each onlooker bee:

- Food source (solution) with better probability value p_m, calculated based on the fitness values provided by the employed bees. The probability of an individual being selected by an onlooker bee is given by [16]:

  $$p_m = \frac{fit_m}{\sum_{m=1}^{SN} fit_m} \tag{9}$$
- Determine new neighboring food source positions using equation (7)
- Calculate the fitness value, fit_m
- Similarly, as in employed bee phase, the best individual is being updated after each iteration based on fitness values.

Scout Bees Phase

The employed bees which cannot improve its solution by using preset number of trials (known as limits or terminating condition) become scout and as such their solutions get discarded. Thus, the scouts get busy looking for new solutions in a random manner and is defined by equation (6) [18].

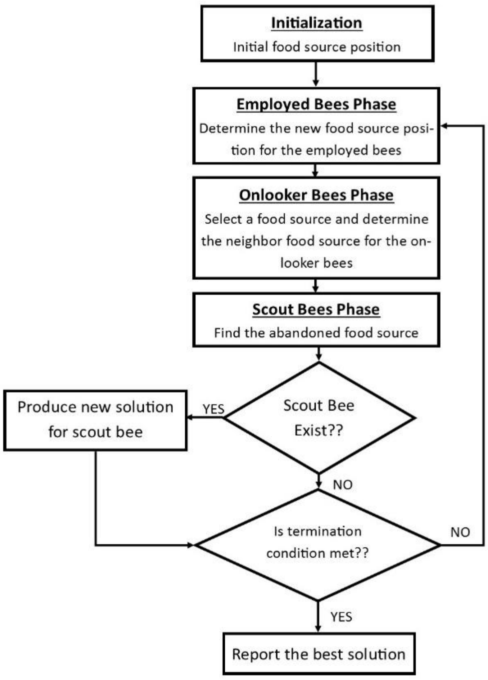


Figure 2: Flowchart demonstrating ABC algorithm

IV. STATE ESTIMATION USING WLS ALGORITHM

Weighted Least square represents an estimation problem that selects the criterion of a solution to the over-defined matrix equations when less states and more measurements are available in the power system [1-2]. Weighted least square have the objective function that minimizes the square of the error. Traditional state estimator utilizes SCADA measurements whose relationship with system states:

$$z_i = h_i(x) + e_i \tag{10}$$

z_i is SCADA measurement vector and its size is $(m\ x\ 1)$

$h_i(x)$ correspond to a nonlinear function vector which relates measurements to states

e_i is the error vector between estimated and measured values of size $(m\ x\ 1)$

Power System State Estimation is a system of overdetermined nonlinear equations and must be solved as unconstrained WLS problem. For weighted least square we need to minimize sum of square of residuals:

$$\min\ f(x) = \sum_{i=1}^{N} W_i^2 \,|\, h_i(x) - z_i \,|^2 \tag{11}$$

$$\min\ f(x) = \frac{1}{2}\left[z - h(x)\right]^T R^{-1}\left[z - h(x)\right] \tag{12}$$

By taking partial derivative of $h(x)$ with respect to state vector x, Jacobian matrix [H] will be obtained.

$$H = \begin{bmatrix} \frac{\partial P_{flow[ij]}}{\partial \theta} & \frac{\partial P_{flow[ij]}}{\partial Vm} \\ \frac{\partial P_{flow[ji]}}{\partial \theta} & \frac{\partial P_{flow[ji]}}{\partial Vm} \\ \frac{\partial P_{inj}}{\partial \theta} & \frac{\partial P_{inj}}{\partial Vm} \\ \frac{\partial Va_i}{\partial \theta} & \frac{\partial Va_i}{\partial Vm} \\ \frac{\partial Q_{flow[ij]}}{\partial \theta} & \frac{\partial Q_{flow[ij]}}{\partial Vm} \\ \frac{\partial Q_{flow[ji]}}{\partial \theta} & \frac{\partial Q_{flow[ji]}}{\partial Vm} \\ \frac{\partial Q_{inj}}{\partial \theta} & \frac{\partial Q_{inj}}{\partial Vm} \\ \frac{\partial Vm_i}{\partial \theta} & \frac{\partial Vm_i}{\partial Vm} \\ \frac{\partial I_{F,Real[ij]}}{\partial \theta} & \frac{\partial I_{F,Real[ij]}}{\partial Vm} \\ \frac{\partial I_{F,Imag[ij]}|}{\partial \theta} & \frac{\partial I_{F,Imag[ij]}|}{\partial Vm} \end{bmatrix}$$

According to Newton method to minimize a function f(x), where $x^{k+1} = x^k + \Delta x$ is

$$\Delta x = \left[f'(x)\right]^{-1} - f\left[(x)\right] = \left[\frac{\partial \nabla_x f(x)}{\partial x}\right]^{-1}\left[-\nabla_x f(x)\right] \tag{8}$$

$$where\, \frac{\partial \nabla_x f(x)}{\partial x} = 2[H]^T [R]^{-1} [H]$$

And the complete formulation for updating states becomes

$$x_{k+1} = x_k + \Delta x = x_k + \left[[H]^T [R]^{-1} [H]\right]^{-1} [H]^T [R]^{-1} \begin{bmatrix} z_1 - h_1(x) \\ z_2 - h_1(x) \\ \cdots \\ \cdots \\ z_m - h_m(x) \end{bmatrix}$$

Δx is the measurement mismatch, which is used as iteration step for next iteration, for Δx to exist non-singularity for gain matrix $[H]^T[R]^{-1}[H]$ is must. R (error covariance matrix of SCADA measurements), iterative procedure terminates when Δx goes below a certain low threshold value.

V. SIMULATION RESULTS

A) IEEE 14 bus system:

As discussed above, simulation is carried out on the IEEE 14 bus test system. In solving the optimization problem in placing PMUs in the proper location, MATLAB based optimization tool CVX is applied for the comparison with the proposed Artificial Bee Colony (ABC) algorithm. Three different optimal solution sets are found from ABC optimization with same number of PMUs (four) which can meet the objective function. The minimum number of PMUs required and the best bus locations for PMU placements are presented in table I below for both the optimization tools.

Table I: Results for Optimal PMU locations

Optimized with	Number of PMUs	PMU locations
CVX	5	2, 6, 7, 9, 13
Artificial Bee Colony (ABC)	4	Set1: 2, 6, 7, 9
	4	Set2: 2, 7, 11, 13
	4	Set3: 2, 7, 10, 13

It can be observed from table I that ABC is giving better results with keeping the number of PMUs only 4 even though they are having some different sets of solutions. CVX optimized the problem with 5 PMUs. In all three sets of solutions given by ABC, the system has become observable.

1) Without any Bad data:

The test case has been made of measurement readings taken from the conventional meters. PMUs are then placed in the obtained optimal locations as presented in table I and state estimation is carried out then to see their impact on the estimation performance. This will lead to find out the optimal solution set. It has been done for each of the solution sets with same redundancy level of measurement values so that a comparison can be made between the solution sets. To make

the test cases, total of 42 measurements are taken with a redundancy level of 1.55. To represent the estimator performance, an indicator has been used which is actually the sum of the differences between all the estimated values and the base case values. Lower value of indicator indicates the better performance of the estimator.
Initially, only the voltage magnitudes and phasors from PMUs are considered to run the estimation. So, for the case of 4 PMUs, we are considering 8 measurement values taken from PMU meters and rest are from conventional SCADA meters.
After that the estimation is carried out with considering the current phasors from PMUs. To keep the redundancy level same, 8 conventional power flow readings are replaced by same amount of current measurements of PMUs. The results obtained are presented in table II below.

Table II: SE indicator with readings from PMUs

Optimized with	PMU locations	Measurement Redundancy	Estimator Indicator	
			Voltage from PMU	Voltage and current from PMU
CVX	2, 6, 7, 9, 13	1.55	3.4668	2.8523
Artificial Bee Colony (ABC)	Set1: 2, 6, 7, 9	1.55	3.4096	2.8256
	Set2: 2, 7, 11, 13	1.55	3.4283	2.8016
	Set3: 2, 7, 10, 13	1.55	**3.3529**	**2.7705**

It is seen from the results that the estimator performance for CVX solution set is not better than ABC even if the PMU numbers are higher. Inclusion of current phasors improved the estimator performance for all the cases as the indicators are much lower when current phasors are included. Results show that the solution set3 for ABC with PMU locations in buses 2, 7, 10, 13 gives the best performance indicator with lowest value.

2) With bad data:

A single bad data is then considered in the measurements series. Power flow of branch 4 to 7 has been taken -0.3526 now which was 0.3526 previously. Because of the presence of this bad data, indicator becomes 8.2919 which was 4.8916 before. Table III clearly indicates that the third solution set of ABC algorithm with PMU locations 2, 7, 10, 13 gives the best solution with lowest indicator among all. All the results are presented below in table III.

Inclusion of current phasors in the place of power flows improved performance as it happened without bad data too. The solution set of CVX, even providing an extra PMU to be installed, could not perform that well in estimation. Placing an extra PMU will not be an economical choice too.

Similarly, like 14 bus, state estimation is then carried out by WLS algorithm after placing PMUs in each of the optimal location sets.

Table III: SE indicator with readings from PMUs after affected with bad data

Optimized with	PMU locations	Measurement Redundancy	Estimator Indicator	
			Voltage from PMU	Voltage and current from PMU
CVX	2, 6, 7, 9, 13	1.55	7.3238	6.4680
Artificial Bee Colony (ABC)	Set1: 2, 6, 7, 9	1.55	5.2336	5.0382
	Set2: 2, 7, 11, 13	1.55	5.2862	5.0217
	Set3: 2, 7, 10, 13	1.55	**5.2014**	**4.9914**

B) IEEE 30 bus System:

The work is then extended to the larger power system with 30 buses. The superiority of ABC over CVX in optimizing problems is being proved already in previous section. Different optimal solution sets by Artificial Bee Colony algorithm for placing PMUs are presented below in table IV.

Table IV: Results for Optimal PMU locations by ABC

Number of PMUs	PMU locations
9	Set1: 3, 5, 8, 10, 12, 18, 24, 25, 29
9	Set2: 1, 7, 8, 10, 12, 19, 23, 26, 29
9	Set3: 1,7, 8, 10, 12, 19, 24, 26, 30
9	Set4: 1, 5, 10,11, 12, 19, 24, 26, 27
9	Set5: 3, 5, 8, 10, 12, 18, 23, 26, 29
9	Set6: 3, 5, 8, 10, 12, 19, 23, 25, 29
9	Set7: 3, 5, 8, 10, 12, 15, 20, 25, 30
9	Set8: 1, 7, 8, 10, 12, 15, 19,25, 29
9	Set9: 1, 5, 10, 11, 12, 19, 23, 25, 27

1) Without any Bad data: Test case has been made with the same redundancy level (1.55) as done for 14 bus system. Results are presented below in table V.

Table V: SE indicator with readings from PMUs

PMU locations	Estimator Indicator	
	Voltage from PMU	Voltage and current from PMU
Set1: 3, 5, 8, 10, 12, 18, 24, 25, 29	12.0423	7.7182
Set2: 1, 7, 8, 10, 12, 19, 23, 26, 29	12.4724	8.0055
Set3: 1,7, 8, 10, 12, 19, 24, 26, 30	12.3997	8.6287
Set4: 1, 5, 10,11, 12, 19, 24, 26, 27	11.9904	8.4069
Set5: 3, 5, 8, 10, 12, 18, 23, 26, 29	12.3182	8.7552
Set6: 3, 5, 8, 10, 12, 19, 23, 25, 29	12.0081	8.2738
Set7: 3, 5, 8, 10, 12, 15, 20, 25, 30	12.0463	8.2971
Set8: 1, 7, 8, 10, 12, 15, 19,25, 29	12.1249	8.2745
Set9: 1, 5, 10, 11, 12, 19, 23, 25, 27	11.7615	8.0818

Only the voltage phasors are considered initially from PMUs. Current phasors are then considered by replacing same amount of power flows to keep the redundancy same. It is seen that inclusion of current phasors improves estimation indicator for all cases and it also changes the optimal solution set. PMU location set 9 was best among all when only voltage phasors were considered. Inclusion of current phasors from PMUs made solution set 1 the best one with lowest indicator 7.7182.

2) With Bad data:

Bad data is then applied in one of the power flow values. Estimator performance deteriorates as expected. Then the PMU measurements are considered as above and the results are presented below in table V.

Table V: SE indicator with readings from PMUs after applying bad data

PMU locations	Estimator Indicator	
	Voltage from PMU	Voltage and current from PMU
Set1: 3, 5, 8, 10, 12, 18, 24, 25, 29	17.7770	14.3027
Set2: 1, 7, 8, 10, 12, 19, 23, 26, 29	18.2584	14.6064
Set3: 1,7, 8, 10, 12, 19, 24, 26, 30	18.1936	15.0873
Set4: 1, 5, 10,11, 12, 19, 24, 26, 27	17.6991	14.8926
Set5: 3, 5, 8, 10, 12, 18, 23, 26, 29	18.0779	15.0889
Set6: 3, 5, 8, 10, 12, 19, 23, 25, 29	17.7401	14.7951
Set7: 3, 5, 8, 10, 12, 15, 20, 25, 30	17.7970	14.8547
Set8: 1, 7, 8, 10, 12, 15, 19,25, 29	17.9509	14.8538
Set9: 1, 5, 10, 11, 12, 19, 23, 25, 27	17.4219	14.6587

The optimal solution sets are similar as the cases of without applying bad measurement. Voltage measurements only from PMUs give optimal solution set 9 as the best solution with least indicator of 17.4219. Incorporation of PMU current measurements results solution set 01 as the best optimal locations as it gives the minimum indicator 7.7182.

VI. CONCLUSION

This paper focused on two main issues: to place PMUs optimally which will not only reduce the installation cost but also make the overall system observable and to check the performance of the state estimator if the PMUs are placed in the resultant optimal positions irrespective of the presence of bad data. It has been found that Artificial Bee Colony based optimal position provides the best estimation performance. For the case of running estimation, redundancy has been taken considerably above 1 which will allow the estimator to work even if some meters fail to provide service. Besides of PMU readings, some measurement readings has been taken from conventional meters with some sort of normal noises. The detail formulation of optimization problem has been presented as well as the formulation of the estimator. It has been observed that the performance of the estimation process was improved when using ABC by indicating the optimal placement of PMUs in the system. The results carried out using the IEEE 14-bus and 30 bus system.

REFERENCES

[1] A. Abur and A. G. Exposito, “Power system state estimation: theory and implementation,” CRC Press, 2004.

[2] A. Monticelli, “Electric power system state estimation,” Proc. IEEE, vol. 88, no. 2, pp. 262–282, Feb. 2000.

[3] Ali Abur, “Power System State Estimation: Theory and Implementation,” 2004.

[4] A Phadke, J Thorp, “History and applications of phasor measurements”, In: IEEE PES power systems conference and exposition, 2006 (PSCE’06):IEEE; 2006. p.331–5.

[5] Lin Y-H, Liu C-W, Chen C-S, “A new PMU-based fault detection/location technique for transmission lines with consideration of arcing fault discrimination-Part I: Theory and algorithms”, IEEE Trans Power Delivery 2004;19:1587–93.

[6] Liu C-W, Thorp JS, “New methods for computing power system dynamic response for real-time transient stability prediction”, IEEE Trans Circuits Syst I: Fundam Theor Appl 2000;47:324–37.

[7] Skok S, Ivankovic I, Cerina Z, “Applications based on PMU technology for improved power system utilization”, In: IEEE power engineering society general meeting, 2007: IEEE; 2007. p. 1–8.

[8] P Yang, Z Tan, A Wiesel, A Nehorai, “power system state estimation using pmus with imperfect synchronization”, IEEE Transactions on Power Systems, Volume:28 Issue:4, 2013.

[9] F Aminifar, M Fotuhi-firuzabad, A Safdarian, A Davoudi, M Shahidehpour, “Synchrophasor Measurement Technology in Power Systems: Panorama and State-of-the-Art,” IEEE transactions, 10.1109/Access, 2015.

[10] Nikolas M. Manousakis, George N Korres, Pavlos S Georgilakis, “Taxonomy of PMU placement methodologies”, IEEE transactions on powers systems, Vol 27, No 2, May 2012.

[11] M. Nazari-Heris, B. Mohammadi-Ivatloo, “Application of heuristic algorithms to optimal PMU placement in electric power systems: An updated review”, Renewable and Sustainable Energy Reviews 50 (2015) 214–228.

[12] Bei Gou, “Generalized Integer Linear Programming Formulation for Optimal PMU Placement”, IEEE transactions on powers systems, Vol 23, No 3, August 2008.

[13] M Gol, A Abur, “Optimal PMU placement for state estimation robustness”, Innovative Smart Grid Technologies Conference Europe (ISGT-Europe), 2014 IEEE PES.

[14] Grant, Michael, Stephen Boyd, and Yinyu Ye. "CVX: Matlab software for disciplined convex programming." (2008).

[15] Rana, Madhurima, Saurabh Bilgaiyan, and Utsav Kar,"A study on load balancing in cloud computing environment using evolutionary and swarm based algorithms", International Conference on Control Instrumentation Communication and Computational Technologies (ICCICCT), 2014.

[16] Murugan, R. and Mohan, M. R.. "Artificial Bee Colony Optimization For The Combined Heat And Power Economic Dispatch Problem", Journal of Engineering & Applied Sciences, 2012.

[17] Bahriye Akay. "Parameter Tuning for the Artificial Bee Colony Algorithm", Lecture Notes in Computer Science, 2009

[18] P. Mathiyalagan. "Hybrid enhanced ant colony algorithm and enhanced bee colony algorithm for grid scheduling", International Journal of Grid and Utility Computing, 2011

An Empirical Study on GSM Spectrums in Bangladesh using SDR Technology

Md. Habibur Rahman, Md. Mamunoor Islam
Department of Electrical and Electronic Engineering
Chittagong University of Engineering & Technology, Chittagong-4349, Bangladesh
E-mail address: Habibur333@gmail.com, mcm.mithu@gmail.com

***Abstract*— With the immense advancement in the field of wireless communication in this modern engineering world, Software Defined Radio (SDR) technology has turned out to be an indisputable emerging technology and presents new challenges for communications engineers. The advancement of SDR system has made significant progress in recent years as it becomes a serious substitute of traditional hardware radio architectures where the mathematical procedures are obligatory to decode and process radio signals using analogue circuitry. Recently, computers have turned out to be powerful enough to do the required mathematical calculations using software. So aim of this paper is to demonstrate the study of GSM spectrums in Bangladesh using SDR technology. This approach will provide the observation of uplink and downlink frequency and channel bandwidth used by GSM operators in Bangladesh. An experimental study was conducted with suitable conditions to examine the feasibility and efficiency of the proposed system. The outcome of experimental result is thoroughly examined in this paper.**

Keywords*— *Software defined radio (SDR); RTL-SDR dongle; GSM Spectrum; Uplink frequency; Downlink frequency

I. INTRODUCTION

Conventionally mobile operators use different uplink and downlink frequency bands. To observe and analyze these GSM spectrums, traditional spectrum analyzer can be used which are expensive and specialized hardware setups [1]. Besides, a special receiver for almost each radio communication standard is needed. In many devices, the radio hardware and the decoder hardware are amalgamated on the same board or are not intended to work individually. As a result, using existing radio hardware for an unintended purpose turns out to be problematic or even impossible, which means high costs for specified radio hardware, e.g. for research purposes [2].

In recent years, SDR technology has turn out to be a revolution by bringing much functionality as software with the reduction of the cost of hardware maintenance and up-gradation [3]. RTL-SDR is an extremely low-priced software defined radio based on DVB-T TV (Digital HD TV) USB receiving dongles which has RTL2832U chip in it. In March 2010, Eric Fry, Antti Palosaari and the Osmocom team first discovered this device who were developing their own SDR at that time [4]. From then, several approaches have been made by the researchers all over the world employing this device on their research works. In 2014, a software-defined sensor architecture for large-scale wideband spectrum monitoring system has been proposed (Damian Pfammatter (et al.) where distributed data collection have been done in real-time over the Internet using RTL-SDR [5]. Another approach have been made by Ken Tapping et al. who have presented SDR technology as an alternative to switched radiometers for continuum radio astronomy [6]. A User-Friendly Android-Based Tool for Spectrum-Analysis based on RTL-SDR have been approached by Jens Saalmüller et al. too [7]. In early 2015, a concept of wireless spectrum analyzer in pocket has also been developed using RTL-SDR (Tan Zhang et al.) [8].

In Bangladesh, Different mobile operator uses different frequency bands which are allocated by the government. The uplink and downlink frequency ranges used by various operators is given below.

TABLE.I ALLOCATION OF GSM BAND IN BANGLADESH [14]

Operator Name	Uplink Frequency (MHz)	Downlink Frequency (MHz)
Airtel	880.5-885.5	925.5-930.5
Teletalk	890-895.2	935-940.2
Banglalink	895.2-900.2	940.2-945.2
Robi	900.2-907.6	945.2-952.6
GrameenPhone	907.6-915	952.6-960

So in this paper, an observation of GSM frequency bands of different mobile operators has been conducted and different parameters such as channel bandwidth, SNR(Signal to Noise Ratio) has been analyzed using RTL-SDR which can be used as an alternative approach of a spectrum analyzing unit with advanced radio capabilities such as wideband tuning and waterfall displays.

II. SDR TECHNOLOGY

A software-defined radio system (SDR) is a radio communication system where components which are typically implemented in hardware such as mixers, filters, amplifiers, modulators/demodulators, detectors etc. are instead implemented by means of software on a personal computer or embedded system. The SDR technology allows a flexible usage of one single device for various signals which differ in bandwidth, frequency, and modulation mode. When using SDR devices, the properties of the devices can be adjusted from within software instead of exchanging hardware components like in traditional radio designs. This has led to advanced radios that previously required complicated analogue hardware now being able to be implemented simply in software. This has reduced the cost of advanced radio capabilities such as wideband tuning and waterfall displays [4]. SDRs are therefore more suitable for performing signal analysis.

One more feature of SDRs is the transfer of a complete signal spectrum in a selected frequency range with a defined sample rate to the computer which means all received data are available in a raw format and can be used without the restrictions and information losses of traditional radio hardware such as caused by a fixed filter bandwidth or signal demodulation. Therefore, one single device can work as a receiver for very different types of signals [2]. Major advantages of SDR technology are described below:

- Provides a very low cost radio.
- Takes up little physical space (Portability).
- Software variety and operating flexibility.
- Wide range of radio spectrum.

III. System Architecture

The projected approach is based on the RTL-SDR device, a multi-purpose wide band software defined radio consisting of economical hardware entity for signal reception and a software portion for signal processing. The hardware part which is available in the form of DVB-T USB dongle, consists of an antenna connected to a tuner chip which is connected to the RTL2832U chip via I2C [7]. The tuner IC has been used for receiving the analog signal and filtering out the desired frequency. Then it transforms this frequency down to an intermediate frequency (IF) generating inphase and quadrature

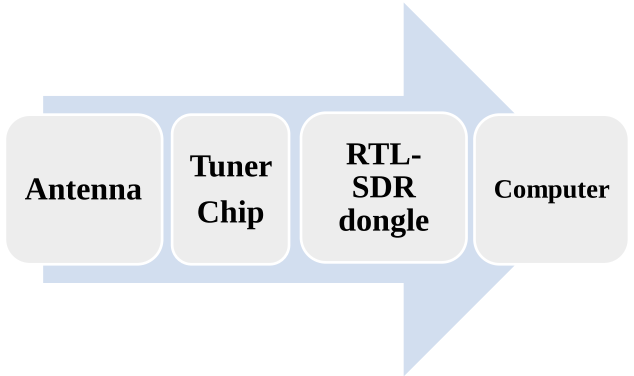


Fig.1. Basic operational Method

components (I/Q signals) and sending them into the RTL2832U chip. This chip then samples the signal with a maximum sampling rate of 3.2 MS/s with 8 bit I/Q samples output. These samples are then sent to the computer via USB. The software part finally processes the raw samples data and illustrates the signal spectrum with waterfall display.

IV. Hardware of The System

A. *The RTL2832U Demodulator*

The RTL2832U is a baseband demodulator which is precisely designed for receiving DVB-T and radio broadcasting. But the application of this demodulator is not limited to these operations. The RTL2832U supports Zero-IF and low IF frequency and has a maximum sample rate of 3.2 MS/s. It receives the IF I/Q signals from the analog tuner IC and outputs the 8 bit I/Q samples [7]. The RTL2832U contains ADC (Analog-to-Digital Converter) and DSP (Digital Signal Processor). It performs DDC (Digital Down-Conversion) via I/Q mixers (phase is 90 degrees apart), digital low-pass filtering, me /Q resampling, and sends 8-bit I/Q data via the USB port [9].

The RTL2832U contains USB 2.0 interface which supports full and high speed modes. This interface has been used to transfer the samples via bulk transfer to the connected host and also to configure the chip through control transfer messages. Another feature of this interface is that it can act as a repeater for the I2C bus. If the repeater is enabled, control messages which is received over USB are forwarded to the I2C bus as well as messages received on the I2C bus are forwarded to the USB port. This mode permits configuration of the tuner chip through USB interface, as the tuner chip is connected to the RTL2832U via the I2C interface.

B. *Tuner Chip*

Almost any DVB-T dongle with the RTL2832U chip can be used with the RTL-SDR drivers. However, one must pay attention to the tuner chip used in the dongle. The tuner chip defines the frequency range of the dongle. There are two commonly used tuners such as R820T and E4000 chips. There are also the less common FC0013 and FC0012. Recently there is also the R828D and FC2580 which are even less common.

TABLE.II Frequency Range of Tuner Chips [4]

Tuner IC	**Minimum Frequency (MHz)**	**Maximum Frequency (MHz)**
R820T	24	1766
E4000	52	2200
FC0012	22	948.6
FC0013	22	1100

Each tuner offers different frequency ranges, gains, amplifiers and filters. But they support a frequency range of at least 60 to 1100 MHz in common which is very wide spread range. At first, the received RF signal is passed into a low-noise amplifier (LNA) where the signal is amplified either automatically or by a manually configurable gain. Next, a certain frequency range is filtered out according to the selected frequency band (VHF II, VHF III, UHF or L-band). After that the mixer transforms the signal into a low frequency IF or Zero-IF and transfers it to the intermediary frequency filter section and gain section where the frequency range is narrowed down to extract the preferred frequency and bandwidth [7]. Fig. 2 illustrates a flowchart of the signal processing inside the tuner IC.

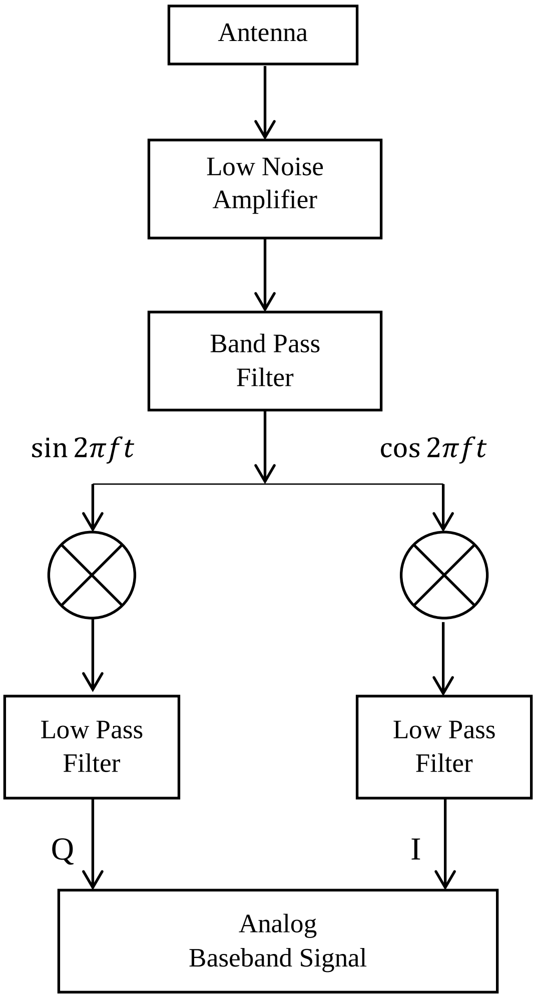


Fig.2. General signal processing inside a tuner IC [7]

C. *Antenna*

In this project, a standard antenna shipped for the RTL-SDR dongles is used. It is a simplistic omnidirectional antenna. Better antenna would certainly result in better reception quality for covering the wide range of frequencies our RF receiver supports.

Fig.3. Hardware of RTL-SDR dongle with Antenna

V. SOFTWARE

More than hundreds software are used on RTL SDR for different purpose in different platform [10]. The most commonly use and available package are given below.

1) *Windows based:*
   a) *Free:* SDR#, HDSDR, SDR-RADIO.COM V2, Linrad, CubicSDR, cuSDR, PowerSDR, QtRadio.
   b) *Paid or trail:* Matlab, Studio1and Sodira.
2) *Linux baseband:*
   a) *Free:* GNU Radio, Linrad, GQRX, QtRadio, CubicSDR, ShinySDR (web based), WebRadio, Multimodes drangelove.
3) *Android-Based:*
   a) *Free:* RFAnalyzer.
   b) *Trail or Paid:* SDR Touch, Wavesink Plus.
4) *Mac baseband:*
   a) *Free:* Linrad, GQRX, and CubicSDR.

From those packages, SDR# is chosen for several facilities. At present it is the most popular windows based free RTL-SDR compatible software. Set up procedure is relatively easy with respect to other one [11]. It has abundant amount of GUI which make it easy to use. It has some advanced features such as different plugins. Though most of plugins are in 3rd party, those are effective.

MATLAB also released the RTL SDR plugging on their R2013b version [12]. With this support package, MATLAB can interface with the RTL-SDR and digital signal processing algorithm can then be written in MATLAB. GNU Radio is another powerful tool for SDR technology [13]. But complexity may arise at the time of installation. Both two software are most powerful in research sector. In recent version of LAB View, there have also scope for interfacing with the RTL-SDR.

VI. Implementation And Analysis

In this project, SDR# has been used as software for more suitable conditions. So, all the experimented results will be shown using SDR# software.

The GSM signal has been received by this RTL-SDR based spectrum analyzing system. Banglalink, a local mobile operator in Bangladesh uses 895.2-900.2 MHz frequency band for uplink and 940.2-945.2 MHz frequency band for downlink and Teletalk, also a local operator uses 935-940.2 MHz band for downlink [14]. These uplink and downlink frequencies received by RTL-SDR are illustrated in Fig.4, Fig.5 and Fig.6.

A. Banglalink Downlink Frequency

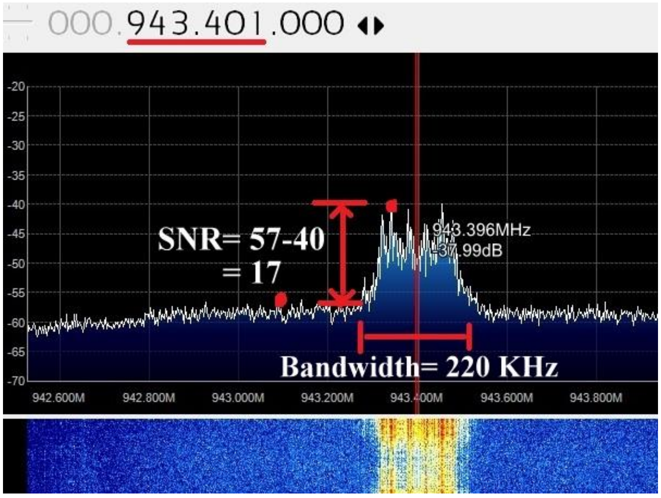


Fig.4. FFT Spectrum of Downlink Frequency with Waterfall Display

B. Teletalk Downlink Frequency

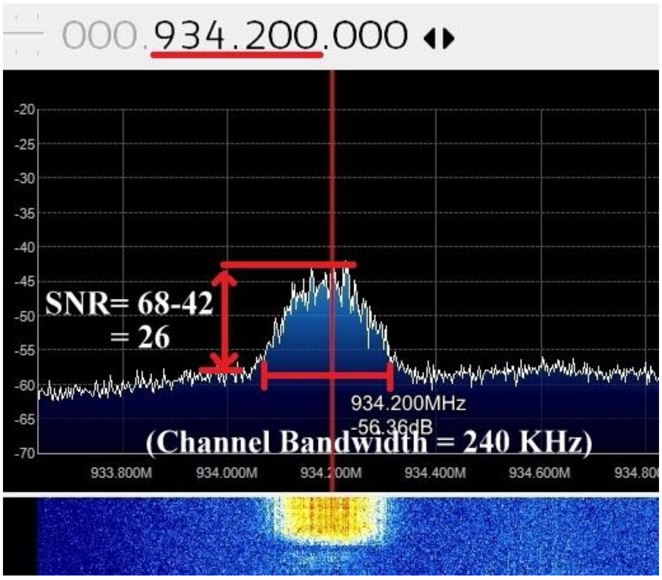


Fig.5. FFT Spectrum of Downlink Frequency with Waterfall Display

C. Banglalink Uplink Frequency

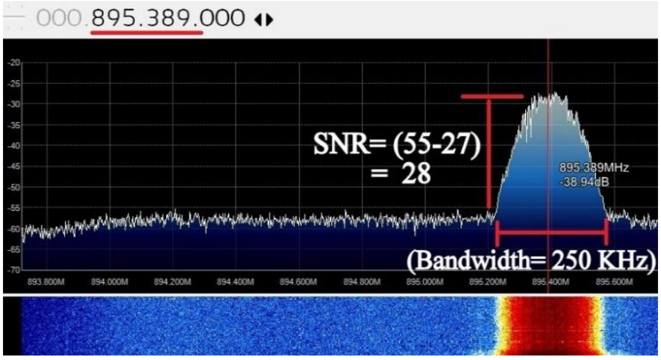


Fig.6. FFT Spectrum of Uplink Frequency with Waterfall Display

TABLE.III Experimented Data

Signal	Measured Approximate Channel Bandwidth (KHz)	SNR (Signal to Noise Ratio)
GSM Uplink (Banglalink)	250	28
GSM Downlink (Banglalink)	220	17
GSM Downlink (Teletalk)	240	26

The experimented data of GSM channel bandwidth analyzed from the figure is close to real data which is conventionally 200 KHz.

VII. Future Work & Conclusion

In future, this RTL-SDR based system may be used to demonstrate and analyze the different kinds of satellite uplink and downlink signal.

In this work, it is shown that the RTL-SDR device can be used as an alternative for spectrum analyzing purposes and by which different GSM spectrums can be analyzed easily. Though the performance of this system is not fully pertinent, the observation can provide different important information such as channel bandwidth, SNR etc. which can be handy for research purposes. Last of all, SDR technology can be used as a modern spectrum analyzing tool which has wide range of frequency tuning conveniences for analyzing the spectrum more proficiently.

References

[1] Ana Nika, Zengbin Zhang, Xia Zhou, Ben Y. Zhao and Haitao Zheng, “Towards Commoditized Real-time Spectrum Monitoring”, HotWireless’14, September 11, 2014, Maui, Hawaii, USA.

[2] Thomas Rudolph, “Analyzing Security-related Signals Using Software defined Radio”, Bachelor’s Thesis. February 12, 2013.

[3] Govarthanam K S, Abirami M, Kaushik J, “Economical Antenna Reception Design for Software Defined Radio using RTL-SDR”, Proceedings of the Intl. Conf. on Innovative trends in Electronics Communication and Applications 2014, page 185-190.

[4] Carl Laufer, “The Hobbyist's Guide to the RTL-SDR: Really Cheap Software Defined Radio”. Kindle Edition, Published on May 14, 2014.

[5] Damian Pfammatter, Domenico Giustiniano, Vincent Lenders, “A Software-defined Sensor Architecture for Large-scale Wideband Spectrum Monitoring”, 14th International Conference on Information Processing in Sensor Networks (IPSN '15), April 14–16, 2015, Seattle, WA, USA.

[6] Ken Tapping, Marcus Leech, “RTLSDR-based, Software Defined Radio Alternative to Switched Radiometers for Continuum Radio Astronomy”, www.sbrac.org/files/DTP_RX.pdf (Retrieved on February 2016).

[7] Jens Saalmüller, Matthias Kuba, Andreas Oeder, “A User-Friendly Android-Based Tool for 868 MHz RF Traffic- and Spectrum-Analysis”, embedded world 2015, February 24-26, 2015, Nuremberg, Germany.

[8] Tan Zhang, Ashish Patro, Ning Leng, Suman Banerjee, “A Wireless Spectrum Analyzer in Your Pocket”, HotMobile’15, February 12–13, 2015, Santa Fe, New Mexico, USA.

[9] Dr. Phil, “Realtek RTL2832U: The mystery chip at the heart of RTL-SDR”, version 1, published on 2015.

[10] http://sdr.osmocom.org/trac/wiki/rtl-sdr (Retrieved on February 2016).

[11] https://uk.groups.yahoo.com/neo/groups/SDRSharp/info (Retrieved on February 2016).

[12] Communications System Toolbox Support Package for RTL-SDR Radio, www.mathworks.com/matlabcentral/fileexchange/44991-communications-system-toolbox-support-package-for-rtl-sdr-radio (Retrieved on February 2016).

[13] http://gnuradio.org/redmine/projects/gnuradio (Retrieved on August 2015).

[14] http://www.spectrummonitoring.com/frequencies/frequencies3.html#Bangladesh (Retrieved on February 2016).

An Efficient Magic Mirror Using Kinect

Md. Moniruzzaman Monir
Dept. of Computer Science and Engineering
BRAC University, BRACU
Dhaka, Bangladesh
monir0805@gmail.com

Md. Nafis Hasan Siddique
Dept. of Computer Science and Engineering
BRAC University, BRACU
Dhaka, Bangladesh
ns.shamit@gmail.com

Nahyan Ebn Hashem
Dept. of Computer Science and Engineering
BRAC University, BRACU
Dhaka, Bangladesh
nahyan.nick@gmail.com

Afsana Pervin Tanni
Dept. of Computer Science and Engineering
BRAC University, BRACU
Dhaka, Bangladesh
meu2ari@gmail.com

Jia Uddin
Dept. of Computer Science and Engineering
BRAC University, BRACU
Dhaka, Bangladesh
jia.uddin@bracu.ac.bd

***Abstract*— To enhance users shopping experience and to spend less time on queuing for fitting room, this paper presents a virtual mirror model using gesture recognition technique. This allows a person to check how a dress looks like and which color is suitable on a person's body. Moreover, it shows users body measurement when users try on virtual cloths. In the proposed model, we used Microsoft Kinect sensors to track user skeleton movement and depth image. The cloths are simulated using unity engine that will represent an environment for the user like the mirror. In addition, we have developed an algorithm for matching up all the motions between the virtual cloths and the human body.**

Keywords- Kinect for windows; Gesture recognition; Skeleton Tracking; Real time image Processing; Virtual try-on

I. Introduction

In this modern era of revolution everyone is depending more and more on technology. According to this flow of development in technology the common definition of shopping is also changing by time to time. Now the traditional shopping has been replaced by online shopping. Now-a-days online shopping or shopping through web is getting more popular because it is saving huge amount of valuable time of the shoppers and also reducing other hassles. Moreover, online shopping is being accepted widely all over the world. More than 85% of world's population has ordered goods over the internet during the recent years [1]. People are getting more attracted to the online shopping because of its extra features or offers like free home delivery, cash on delivery, and different kinds of discounts. However, it has a significant drawback- this method is not being accepted by all peoples as there is no surety that the delivered goods or cloths will be according to the expectation of the customer. Although customers can find all the description of the cloth like style, size, color fabric and other features through the web page, but they cannot determine whether the cloth is exactly suitable for their own style, color, size and other aspects. Therefore, the delivered clothes might also not fit the customers [2]. Previously, a number of researchers worked on this area to overcome the problems of online shopping. The researchers came up with an idea of virtually try the dresses or clothes so that the user do not have to try it physically [3,4]. Among the works, we have chosen "Magic Mirror Using Kinect," by A. B. Habib, A. Asad, W.B. Omar, BRAC University (2015). In this paper authors proposed a model, which has the following limitations:

- User needs to move to adjust the cloth within his or her body.
- Dresses are not accurate to the body shape.
- Use 2D dresses.
- Use no user interface.

To overcome the limitations, we proposed a concept of real time virtual dressing room [3]. As mirrors are indispensable objects in our lives, the capability of simulating a mirror on a computer display or screen, augmented with virtual scenes and objects, opens the door for solving the major drawback in online shopping concept. An interactive Mirror could enable the shoppers to virtually try clothes, dresses using gesture-based interaction [5]. The proposed method has the following features:

- Use a static position.
- Varity of dresses.
- Display size of dresses.
- Use hand gesture and improved user interface.

In this paper, gesture based interaction techniques are used in order to create a virtual mirror for the real-time virtualization of various clothes. Similar to looking into a mirror when trying on new clothes in a shop, we create the same impression but for virtual clothes that the customer can choose individually [6]. For that purpose, we replace the real mirror by a large display that shows the mirrored input of a camera capturing the body skeleton of a person. The use of a hierarchical approach in an image pyramid enables real-time estimation at frame rates of more than 30 frames per second.

This paper is organized as follows- Section II provides a detailed overview of proposed model. Section III includes the experimental setup and outcomes of the proposed implementation. Finally, the paper concludes in Section IV.

II. PROPOSED MODEL

This project mainly focuses on creating a virtual dressing room. This requires real-time tracking of the user skeleton as well as realistic virtual clothing. For the pose tracking Microsoft Kinect sensor is used which gives more complete and accurate tracking of the user pose than the marker based or image feature based tracking which is traditionally used in augmented reality applications. For the clothing we created a set of 3D dress models that can be rendered into the screen. The focus of this project is on realistic interaction and simulation between the user and the virtual clothing.

To achieve this, the clothing needs to:

- Be aligned correctly with the user position and pose.
- Move and fold realistically.
- Be realistically rendered into the environment such as ambient lighting.

Figure 1 shows a detailed block diagram of system implementation of our proposed model.

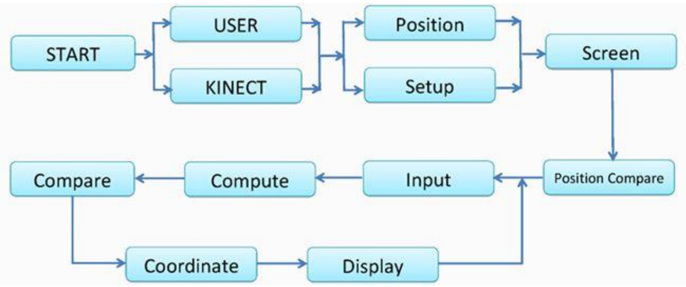


Figure 1: System implementation flow.

A. System implementation flow

1. *User and Kinect*: Human body act as input source and by Kinect we take that input. Both Kinect and raw data from user are taken parallel. After that all information passes to the next session as input.
2. *Position and Screen setup*: After taking raw values from user and Kinect the system measures the position of user and set screen setup.
3. *Screen*: Screen will process the raw data and change it some frame model and will show into the screen.
4. *Position comparison*: In this stage our code will compare each array with previous array of frame to identify the rotation of human.
5. *Input*: At this stage system will save all the raw data for further calculation.
6. *Computation*: The raw data are compiled and also compute the dress model data.
7. *Compare and Coordinate*: From previous step's raw values it will compare human model values with dress model values take some point to coordinate them.
8. *Display*: after doing all this process our system will display the final output. Then it will prepare itself for next input.

Figure 2 illustrates the process flow of our proposed model.

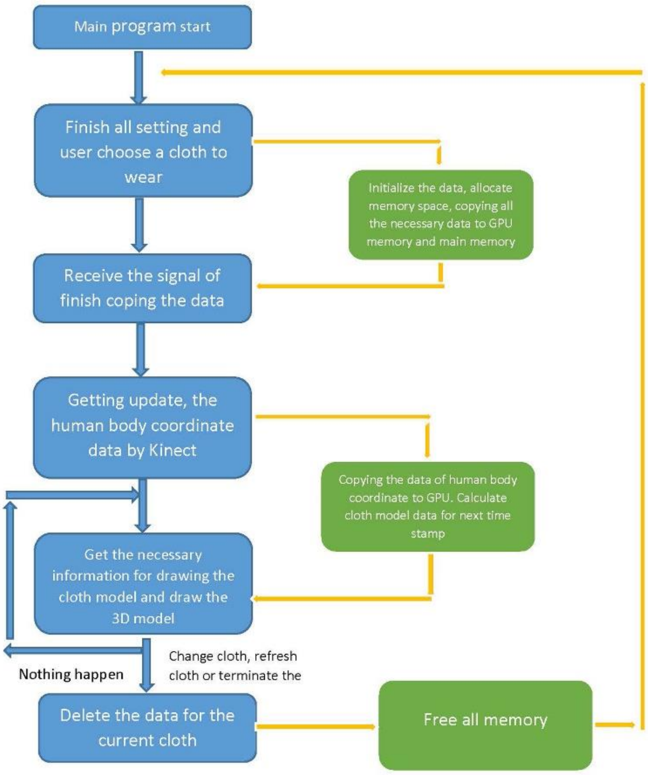


Figure 2: Process flow of proposed model.

B. Process flow

Process flow is the diagram of how the system will run in the software stage. The description of the diagram will help to understand what is happening inside the system from system start to simulation of virtual cloth.

There are some steps that need to be fulfilled to successfully run the complete system. Some of the steps are in loop, as one step needs to be iterating over and over to get the actual result.

First, the starting of the program initiates a series of activities. The memories inside the program are allocated for the different part of the system, where initially the memory is free.

Second, the system gives the user a view of interactive screen. Here the user will get the options to choose the cloth they want to try on. The Kinect sensors get the measurement and the structure of the user.

Third, the system gets the notification or signal that the cloth data for comparing with the user is now available and the other instruction for the system to get the value properly is ready.

In between the second and third step there is a step that would be looped for every cloth. Whenever the cloth data is called this steps are to be executed.

The data is initialized in the fourth step. The memory gets the allocation for the cloth data and the necessary data from the database is copied in the system. The data that the GPU (Graphics Processing Unit) will use to render the cloth is also copied in the allocated main memory and the GPU memory.

Fifth, The Kinect takes action and uses the sensors to get the human body structure and the coordinate of the body. It helps to identify the body structure, positions and coordination of the skeleton point and the depth perception from the Kinect sensor.

Sixth, the necessary data taken from the previous steps helps the system to generate a skeleton from the real images. The skeleton gives the system the base of the virtual dress. The virtual dress is then drawn over this skeleton structure. The Unity Engine gives us the option to apply physics and gravitation in the virtual cloth. So the cloth is drawn based on those options we have in an advance dimension than 2D. When the dress is drawn or generated in the screen we will get an augmented reality based approach of the cloth over the human user.

Seventh, in between the fifth and sixth steps there is more important and crucial steps that loops around for every bit change in the system. Every time when the human body or the user changes the position or moves a slight the data is passed by the sensors to the system. The data is then copied and the coordinate is given to the main memory and the GPU memory. When the data is in GPU, then the system uses the comparison algorithm to identify the new portion and to generate cloth for the change of the body. The new position is used for new cloth generation.

Eighth, to get into this step the user needs to give some specific instructions. If the user wants to change the cloth or the user want to end this session of the system use, and then these steps will be activated. If the user does nothing, then as long as the user is inside the fixed position the cloth will be continued to be generated for every movement change. If the user chooses to change the cloth, then the system will go to the next step. If the system gets the instruction, then the cloth data will be deleted. There will be no data of the user body structure and the cloth data in the memory.

In the last step, the memory will be free for the new data of the cloth. The user instruction for changing the cloth with enable the system for going through this step and clear the current cloth data from the memory. If the session is expired by the user, then again memory will be cleared. This memory clear will get the system in a loop which backs the system to the second step. So this is the main loop for the system to start again for change of cloth, refresh the system or for terminate the system session.

In between second and ninth step this step is very important. When the memory is free, it means that the allocation of the memory in main memory and the GPU memory will be cleared and the data will be removed for the new cloth and to get the human structure again.

III. EXPERIMANTAL SETUP

In this section, total overview and summary of the system is presented, which includes basic experiment setup on. The program flow will be explained briefly. In addition, the design of the user interface debugging functionality is also discussed:

A. Basic setup

1. Microsoft Visual Studio

Microsoft visual studio is an integrated development environment (IDE) from Microsoft. It is an integrated solution which enables the users to develop console and graphical user interface applications along with Windows Forms Applications (WFP), web sites, web applications, web services etc [7]. Visual studio supports almost all kinds of programming languages including built-in languages such as C, C++ [8] (via Visual C++), VB.NET (via Visual Basic .NET), C# (via Visual C#), and F# (as of Visual Studio 2010 [9]). All of these languages are built on top of the .NET Runtime (known as Common Language Runtime or CLR) and produce the same intermediate output in Microsoft Intermediate Language (MSIL) [10].

Like any other IDE, it includes a code editor that supports syntax highlighting and code completion using IntelliSense for variables, functions, methods, loops and LINQ queries [11]. IntelliSense is supported for the included languages, as well as for XML and for Cascading Style Sheets and JavaScript when developing web sites and web applications [12][13]. Autocomplete suggestions appear in a modeless list box over the code editor window, in proximity of the editing cursor. In Visual Studio 2008 onwards, it can be made temporarily semi-transparent to see the code obstructed by it [11]. The code editor is used for all supported languages.

2. Microsoft Kinect

Kinect is a motion sensing input device developed by Microsoft in early 2010 [6]. Kinect is mainly used in Xbox 360 console and for Windows PCS [6]. For Windows, a Kinect sensor consists of an RGB camera which can store up to three channel data of resolution 1280×960. It has an IR (infrared) emitter which emits lights beams and an IR depth sensor that reads the reflected beams and process the information to measure the distance between the object and the sensor. It also has multi-array microphone that can capture sound and detect the location of the source and direction of the audio wave. It has a practical ranging limit of 40cm–3.5m for Windows and the frame rate is 30 FPS (Frames per Second) [14]. Figure 3 depicted a picture of a Kinect.

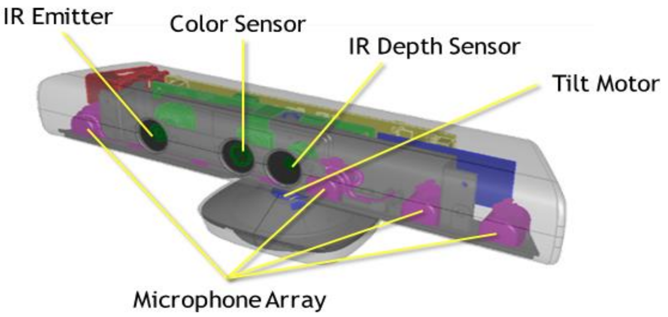


Figure 3: Kinect components (https://msdn.microsoft.com/en-us/library /jj131033.aspx.)

Kinect SDK is developed to enable developers to develop applications in C++, C# or Visual Basic by using Microsoft Visual Studio [4]. It is capable of capturing front body 2D motion, gesture, facial and voice recognition [6], skeletal tracking and advance audio capabilities [15]. To setup virtual mirror we need Kinect sensor to record skeleton and depth data and capture the RGB video stream.

In addition, the software has the capability to recognize and track human body. The software runtime converts depth data into about 20 skeleton joint points of human body to track up to two persons in front of the camera [16]. Figure 4 shows the joint points of a human body detected by Kinect [18].

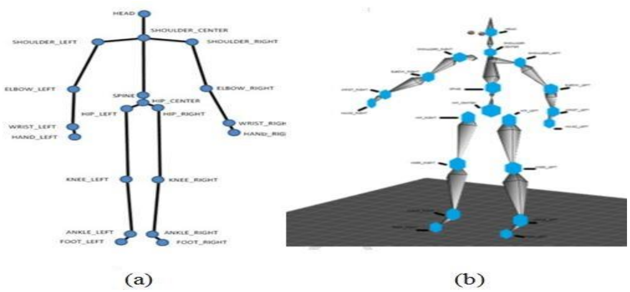


Figure 4: (a) Skeleton joints found by Microsoft Kinect (https://ometechnology.files.wordpress.com/2013/03/durango-kinect.jpg) (b) joint structure on avatar [18]

3. Computation

To virtually set the cloths on the user and display it in real time first we need to detect the skeleton of the user and record both skeleton and depth data for processing. After that we need to define key joint points using skeletal tracking algorithm and compute the right simulation of virtual cloth to display the right 2D cloth. Lastly, we combine real time video and clothing simulation on skeletal points and displayed it.

4. Display/Screen

To give the user a mirrored image impression there is a screen or display in front of the user. Data processing has to be accurate and mirroring the real time video essential in order to give the user an impression of mirror.

Figure 5 shows the application flow of the proposed model.

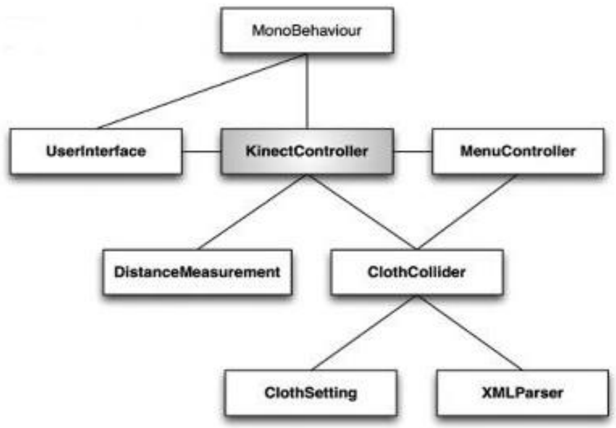


Figure 5: Application Flow of Proposed Model

5. Ordinary Differential Equation(ODE)

For calculating the velocity and the force of the particles in each frame, differentiation is required. The followings are the 3 equations that were chosen for the calculation of differentiations.

Euler Method [19]:

$$y_n = y_n + \text{hf}\,(t_n, y_n)$$

Heun's Method [20]:

$$y'_{n+1} = y_n + \text{hf}\,(t_n, y_n),$$

$$y_{n+1} = y_n + \frac{h}{2}\,(\text{f}\,(t_n, y_n) + \text{f}\,(t_{n+1}, y'_{n+1}))$$

Runge-Kutta Method [21]:

$$y_{n+1} = \frac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4)$$

$$k_1 = \text{hf}\,(t_n + y_n)$$

$$k_2 = \text{hf}\,(t_n + \frac{1}{2}h, y_n + \frac{1}{2}k_1)$$

$$k_3 = \text{hf}\,(t_n + \frac{1}{2}h, y_n + \frac{1}{2}k_2)$$

$$k_4 = \text{hf}\,(t_n + h, y_n + k_3)$$

After testing on the three differentiation method, only Runge-Kutta method is suitable for the program. The other two methods have relatively big error range of the result and will make the cloth become unstable.

B. Full application flow

The user will stand in front of the virtual mirror. Then the user will perform a 180-degree calibration pose so that Kinect could record the 3D depth image, skeleton and joint points for calculating the structure of human body. After executing calibration and recording measurement data, the system will continue. Now by using Kinect's skeletal tracking algorithm, it is now possible to access the skeleton parameters and joint point. Now an algorithm needs to keep the track of the change in key joint points to track movement of the user. If any problem occurs during the calibration process, the system will restart. The virtual mirror (screen) will display user's real time mirrored videos. The user will then choose the cloth that the user wants to see virtually in the mirror. The dynamic cloth will appear over the user's garment. The user will have to maintain a minimum distance to track by the Kinect to get the skeleton points and depth data. After the simulation the user will have to move out of the mirror area or sensor's view–point to reset the whole system.

Figure 6: Kinect setup and human ditection.

C. User interface

As this is a virtual mirror and it can be implemented anywhere from commercial cloth store to personal dressing mirror, there is no need of external display for the application interface. The virtual mirror is used to show the interface and it is controlled by gesture recognition feature of Kinect. Particularly swipe gesture will be implemented and will be distinguished by right and left swipe method. This swipe method can be tracked and implemented by using the Kinect sensor. By using this feature user will be able to choose their desired cloths to display. Figure 7 indicates that the users would have to be at a minimum distance from the Kinect to be able to successfully track by Kinect [1].

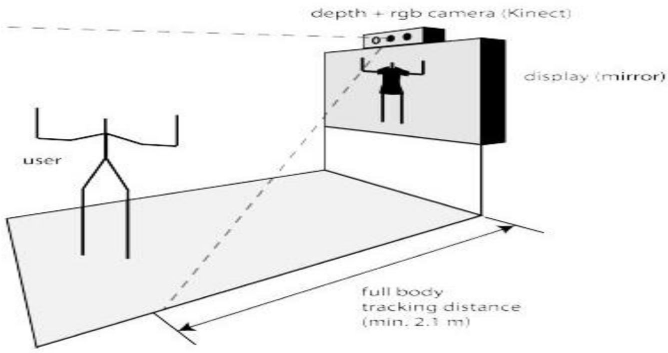


Figure 7: Minimum distance for tracking [1].

D. Debugging:

In the process of debugging the system, we need to get the live value from out testing. The debugging process can be differentiate into two steps:

1. Checking value.
2. test.

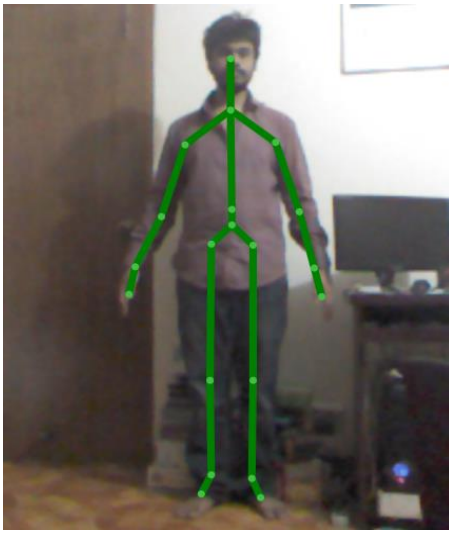
- Checking Values

In the debugging process, we first need to compile and create the environment of the test to check if we are getting the proper skeleton value and the RGB images. We need these values to do the human tracking that will enable us to create dynamically generated clothes for our system. The values that we would get by calculating the distance of every person's body structure from the skeleton and real images will help us to determine our cloth structure and the stretch point of our cloths. Using these values, we can determine the efficiency of our system.

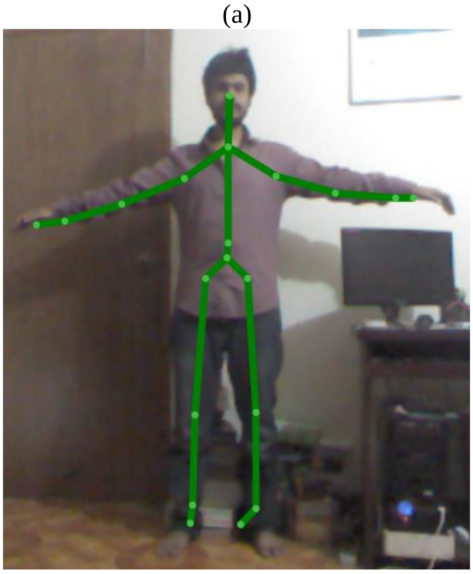
- Testing

The testing of the code and the system is another important part. We need to test the system using live input to detect the movement and then apply the virtual cloth simulation. From the testing it and we will be able to identify the proper simulation and error in the system. As we have implemented and designed the proper system with minimum error hence the testing will give us the sufficient data to analyze and calculate our virtual mirror effectiveness. In the experimental part,

we have generated a skeleton and real image based compound display to identify the movement. There is some of our experimental images that show that we have successfully implemented the code and we can track the movement of the human body. Currently this is a 2D structure based model where we can implement 2D cloths and we will develop it in 3D structure model so that we can implement our main goal of 3D cloth simulation and the virtual mirror. In addition, detection of skeletal joint points are shown in Fegure7.



(a)



(b)

Figure 8: Body Simulation and Tracking with 2D cloths.

IV. CONCLUSION

In this paper, we introduce a virtual dressing room application where avatar and cloth generation, real time tracking technologies up to an overview of comparable virtual try-ons. Subsequently a closer look on the technologies and frameworks that were used for the implementation of the virtual dressing room was taken. After this the different aspects of the design process up to the construction of the garment models was highlighted. This is followed by the implementation, describing the cloth colliders and the behavior of the garment, for instance. In the last section the tests were executed, also discussing the output, the appearance and the interaction with the virtual dressing room. Overall, the presented virtual dressing room seems to be a good solution for a quick, easy and accurate try-on of garment. The Microsoft Kinect offers the optimal technology for a successful implementation. Compared to other technologies like augmented reality markers or real-time motion capturing techniques no expensive configurations and time-consuming build-ups are required. From this point of view, it is an optimal addition for a cloth store. A simple setup of the system can also be assembled at home since the minimum requirements are a computer with a screen and a Kinect.

REFERENCES

[1] U. Cheema, M. Rizwan, R. Jalal, F. Durrani, N. Sohail, "The Trend pf online shopping in 21st century: Impact of enjoyment in tam model" Asian Journal of Empirical Research, vol. 3, no. 2, pp. 131-141.

[2] L. Zhao, J. Zhou, "Analysis on the Advantages and Disadvantages of clothing Networking Marketing" International Journal of Business and social science, vol. 6, no. 4(1), pp.147-151, (2015).

[3] A.B. Habib, A. Asad, W.B. Omar, "Magic Mirror Using Kinect," BRAC University (2015).

[4] S. Giovanmi, Y.C. Choi, J. Huang, E.T. Khoo, K. Yin, "Virtual try-on using Kinect and HD Camera," MIG-2012, vol. 7660, pp. 55-65, (2012).

[5] P. Presle, "A Virtual Dressing Room based on Depth Data," Vienna Uninversity of Technology, Klosterneuburg, pp. 25-36, (2012).

[6] D. Ravì. Kinect: the next generation of motion control. Retrieved from http://www.dmi.unict.it/~battiato/CVision1112/Kinect.pdf.

[7] H. P. Halvorsen, (2014, March 12). Introduction to Visual Studio and C#. Retrieved from http://home.hit.no/~hansha/documents/microsoft.net/tutorials/introduction%20to%20visual%20studio/Introduction%20to%20Visual%20Studio%20and%20CSharp.pdf.

[8] Brenner, Pat (19 July 2013). "C99 library support in Visual Studio 2013", Visual C++ Team Blog. Microsoft.

[9] D. Chai, and K. N. Ngan, Face Segmentation using Skin-Color Map in Videophone Applications, IEEE Transactions on Circuits and Systems for Video Technology, vol. 9, no. 4, pp. 551-559, (1999).

[10] A. Sur, Visual Studio 2012 and .NET 4.5 Expert Development Cookbook, vol.1, Chapter No. 1 "Introduction to Visual Studio IDE Features, "(2013).

[11] Guthrie, Scott. "Nice VS 2008 Code Editing Improvements", July 28, 2007.

[12] Guthrie, Scott. "VS 2008 JavaScript IntelliSense," June 22, 2007.

[13] Guthrie, Scott. "VS 2008 Web Designer and CSS Support," July 25, 2007.

[14] Kinect for Windows Sensor Components and Specifications. Retrieved from https://msdn.microsoft.com/en-us/library/jj131033.aspx.

[15] H. Fairhead, "All About Kinect,". Retrieved from http://www.i-programmer.info/babbages-bag/2003-kinect-the-technology-.html.

[16] G. Yolcu, S. Kazan, and C. Oz, "Real Time Virtual Mirror Using Kinect," Baikan journal of Electrical & Computer Engineering, vol. 2, no. 2, pp. 75-78, (2014).

[17] A. Kar, "Skeletal Tracking using Microsoft Kinect the Microsoft Kinect sensor," Methodology (2010).

[18] M. Kotan, and C. Oz, "Virtual Mirror with Virtual Human using Kinect Sensor," 2nd International Symposium on Innovative Technologies in Engineering and Science, Karabuk University, Turkey, pp. 730-738, (2014).

[19] Paul's Online Math Note. Retrived from http://tutorial.math.lamar.edu/Classes/DE/EulersMethod.aspx

[20] Department of mathmatics, University of Kansas State. Retrived from https://www.math.ksu.edu/math240/book/chap1/numerical.php.

[21] Erik Neumann, 2004-2010. Retrived from http://www.myphysicslab.com/runge_kutta.html

Blind Image Restoration by using PCA-Subspace Generation and Image Quality Optimization

Brian Sumali[1], Haslina Sarkan[2], Nozomu Hamada[1]
MJIIT[1], AIS[2]
Universiti Teknologi Malaysia
Kuala Lumpur, Malaysia
sbrian2@live.utm.my, (haslinams, hamada)@utm.my

Yasue Mitsukura
Department of System Design Engineering
Faculty of Science and Technology, Keio University
Yokohama, Japan
mitsukura@sd.keio.ac.jp

***Abstract*—Image blurring process is commonly formulated as two-dimensional convolution between the latent image and the blurring system. Blind image restoration problem is to estimate the latent image only from the blurred image. Conventional blind image restoration techniques tend to solve this problem by estimating the blurring system and therefore their effectiveness are dependent to the accuracy of their estimation. Principal Component Analysis (PCA)-based restoration technique, however, do not employ PSF estimation and still gives high restoration quality. PCA-based techniques have two different roots. The first is to boost the high-frequency component lost during the blurring process by maximizing the image variance. The other comes from source-separation using PCA. Previously we proposed a PCA-generated subspace for blind restoration and proved its superiority to conventional methods. However, the algorithm should be improved. This study proposes a sign-determination method and modify the image quality optimization technique. The noise robustness and the application to other blurs are also investigated in this paper. From the experiments, we can see that the proposed method gives higher restoration quality both for simulated blur images and real images than conventional methods.**

***Keywords*—Blind image restoration; Single image restoration; Principal Component Analysis; Image enhancement; Image quality**

I. Introduction

Image restoration is to remove all degradations from a corrupted observation and estimate the original, pristine images. Degradations come in many types such as motion blur, camera defocus, noise, atmospheric turbulence, and many others. Typical image restoration techniques assume blur-induced degradations as a system called point-spread-function (PSF). The corrupted observation is assumed as a result of a convolution between the pristine image with the PSF and addition with noise. Based on these assumptions, typical image restoration techniques estimate the inverse of degradation system and the restoration is performed by convolving the corrupted observation with the inverse system. Therefore, an image restoration is a deconvolution process. Although these assumptions imply that the degradation process is a linear process and shift-invariant, which is not true for real world's cases, this model is still used because it simplifies the problem. The simplest image restoration algorithm is the inverse filter [1].

In real world cases, there is almost no information available regarding the PSF nor the original image and it is usually defined as an ill-posed problem where there are multiple possible combinations of latent image and PSF. The techniques designed to tackle this ill-posed problem is called blind image restoration or blind image deconvolution technique. Up to now, numerous blind deconvolution algorithms have been developed in many fields, such as medical imaging, astronomy, remote sensing, optics and daily life. An overview of conventional blind image restoration techniques are available from [2]-[5].

Conventional blind image restoration techniques can be categorized into two types: The first kind is by estimating the PSF of the degradation system from given observation and perform deconvolution using the inverse system. The algorithms in this category are usually straight forward and only concerned about PSF estimation. The algorithms on the second category are called joint estimation which simultaneously estimate the PSF and the latent image. These algorithms are usually more complex than the first category.

Lane and Bates proposed zero sheet approach [6] for multi-dimensional deconvolution. Ayers and Dainty [7] applies alternating estimation between the PSF and the original image in the Fourier domain and under the non-negativity constraints during iterations. Another example is the blind implementation of Lucy-Richardson algorithm [8],[9] that estimates both PSF and original image using Lucy-Richardson Algorithm by Fish et al. [10]. While not needing correct estimation of the degrading PSF, this algorithm still imposes prior assumption for PSF.

Gradient magnitude distribution of natural images were incorporated for camera-shake or motion-blur PSF estimation by Fergus et al. [11]. However, their algorithm still impose a user input and not fully-automatic. Recently, this weakness was improved by L. W. Chang and J. T. Chang for automated patch selection [12]. Levin et al. [13], [14] proposed a MAP_k algorithm. They argued that MAP estimation over the blurring system is more accurate compared to conventional $\mathrm{MAP}_{x,k}$ algorithms [11], [15], [16] that often failed to perform restoration.

In contrast with conventional blind restoration algorithms, PCA-based restoration algorithms do not use inverse filtering to recover the image. This approach was pioneered by Li et al. [17]. Blurring process reduces the image's variance and removes the high-frequency component. Their idea is to boost the missing high-frequency component by enhancing the variance of the deconvolved image. They designed a deconvolution FIR filter with unity filter norm constraints by maximizing variance of output image of the deconvolution filter. Their succeeding work [18] improves their previously multi-frame algorithm to solve single-channel algorithm by creating ensemble of horizontal and vertically shifted degraded images. Both approaches perform the restoration by adding the first principal component to the degraded image.

Our previous approaches [19]-[22] implemented PCA but with a several differences from [18]. Ensemble generation process by Gaussian smoothing instead of shifting the observed image was proposed in [19]. The restoration quality was also improved by incorporating the second principal components instead of only the first. NIQE [23], a no-reference image quality assessment index (NR-IQA) was introduced in [20] to control the iteration number.

In [21] we proposed to employ magnitude scaling to the principal components and also investigated the optimal parameters for ensemble generation process. We generalized the PCA-based blind image restoration into an optimization problem by maximizing a NR-IQA's score in [22]

Although this study is also focused on atmospheric turbulence-degraded image and based on PCA for restoring high-frequency components in the same way as our previous studies [21], [22], some investigations and modifications have been done to improve the previous approach. The novel contributions are as follows:

1. We introduced sign determination and modified the optimization algorithm. This improves the restoration quality.
2. The noise robustness of the proposed algorithm is evaluated and improved.
3. We show that our method can be applied for other types of blur, not only Gaussian blur or atmospheric turbulence.

The rest of this paper is structured as follows: In II, a review for our PCA-based image restoration methods is presented. Our proposed modifications are described in III, and some experimental results are shown in IV. The paper is then concluded in V.

II. PCA-based Image Restoration

A. Basic idea of PCA-based approach

Fig. 1 illustrates a basic idea behind our PCA approach. The original image and the blurred image are represented by two vectors $f, g(= g_1)$ in image vector space respectively. Another set of image vectors $g_2, g_3, \cdots, g_M$ are generated by further blurs with known PSF where g_{i+1} is blurred from g_i recursively up to $i = M - 1$. By utilizing the image vector which is derived from subtracting image g_2 from g_1, we may generate a 1-D subspace or linear space with scalar parameter λ as;

$$g_1 + \lambda(g_1 - g_2) \tag{1}$$

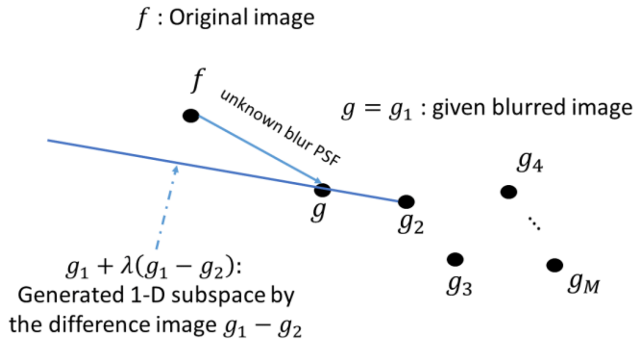


Fig. 1 Basic idea of PCA-based image restoration

Since g_2 is the blurred image from g_1, the subtraction part $\lambda(g_1 - g_2)$ with appropriate λ would be considered as an approximated high frequency component image which was lost by the further blurring process. Conventional unsharp masking [24] uses a similar idea which was developed for image enhancing such as edge or contrast enhancement. However in the PCA-based technique we generate multiple blurred images and apply PCA for generating higher orthogonal subspace. If we generate M blurred image and utilize the resulting major J orthogonal principal components as shown in Fig. 1, restoration process will be performed in J-dimensional subspace. A detailed mathematical development about the proposed approach will be discussed the proceeding sections.

B. PCA-subspace image restoration

Before applying the restoration method, the blurred image g needs to be preprocessed into column vector $\boldsymbol{g}$. The restoration method we proposed consisted of several steps:

1. Generate an ensemble of M blurred images.

$$\begin{gathered} \boldsymbol{g}_1 = \boldsymbol{g} \\ \boldsymbol{g}_i = \boldsymbol{g}_{i-1} * b(\sigma);\ i = 2\sim\mathrm{M} \end{gathered} \tag{2}$$

where $\boldsymbol{g}$ is the blurred image and $b(\sigma)$ is the Gaussian smoothing PSF with standard deviaton of σ.

2. Compute the average image $\boldsymbol{\Psi}$ from the ensemble.

$$\boldsymbol{\Psi} = \frac{1}{M}\sum_{i=1}^{M} \boldsymbol{g}_i \tag{3}$$

3. Center the ensemble by removing the mean.

$$\begin{gathered} \boldsymbol{\phi}_i = \boldsymbol{g}_i - \boldsymbol{\Psi} \\ \boldsymbol{A} = [\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \dots, \boldsymbol{\phi}_M] \end{gathered} \tag{4}$$

4. Extract the principal components

The principal components from the ensemble is the eigenvector of covariance matrix:

$$\boldsymbol{C}_\phi = \frac{1}{M}\boldsymbol{A}\boldsymbol{A}^T \tag{5}$$

where the size of $\boldsymbol{C_\phi}$ is $(R \cdot C) \times (R \cdot C)$ in case of $R \times C$ being the size of $\boldsymbol{g}$. Computing $\boldsymbol{C_\phi}$ itself needs extremely large resource. To reduce the computational burden, Turk et al. [ref from nakamura] proposed to compute the eigen-decomposition of $\boldsymbol{A}^T\boldsymbol{A}$ first.

$$\begin{aligned} \boldsymbol{C}_{\boldsymbol{\phi}}^T &= \boldsymbol{A}^T\boldsymbol{A} \\ \boldsymbol{C}_{\boldsymbol{\phi}}^T\boldsymbol{u} &= \lambda\,\boldsymbol{u} \end{aligned} \tag{6}$$

Then, w pre-multiply both sides with $\boldsymbol{A}$, to get the principal components of $\boldsymbol{C_\phi}$.

$$\begin{aligned} \boldsymbol{A}\boldsymbol{C}_{\boldsymbol{\phi}}^T\boldsymbol{u} &= \lambda\,\boldsymbol{A}\boldsymbol{u} \\ \boldsymbol{A}\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{u} &= \lambda\,\boldsymbol{A}\boldsymbol{u} \\ \boldsymbol{C}_{\boldsymbol{\phi}}\boldsymbol{A}\boldsymbol{u} &= \lambda\,\boldsymbol{A}\boldsymbol{u} \\ \boldsymbol{v} &= \boldsymbol{A}\boldsymbol{u} \end{aligned} \tag{7}$$

where $\boldsymbol{v}$ is the principal components of $\boldsymbol{C_\phi}$

5. Estimate the pristine image

We define the restored image as a linear combination of scaled principal components.

$$\widehat{\boldsymbol{f}} = \widehat{\boldsymbol{f}}_J = \boldsymbol{\Psi} + \sum_{\boldsymbol{i=1}}^{\boldsymbol{J}} \alpha_i\,\boldsymbol{v}_i;\ J < M, J \in \mathbb{N} \tag{8}$$

where $\boldsymbol{v_i}$ is the i-th principal components and α_i is the "weight", a scalar to control the magnitude of $\boldsymbol{v_i}$. The set of optimal weights $\alpha_i(i = 1{\sim}J)$ are successively determined by maximizing the adopted NR-IQA value of the estimate $\widehat{\boldsymbol{f}}_J$. As an illustration, the most ideal restored image $\widehat{\boldsymbol{f}}$ will be the projection of latent image $\boldsymbol{f}$ to the subspace spanned by principal components $\boldsymbol{v_1}{\sim}\boldsymbol{v_J}$.

Optimal α_1:

$$\widehat{\alpha_1} = \arg\max\ NRIQA(\widehat{\boldsymbol{f}_1}) \tag{9}$$

where

$$\widehat{\boldsymbol{f}_0} = \boldsymbol{\Psi} \tag{10}$$

$$\widehat{\boldsymbol{f}_\iota} = \widehat{\boldsymbol{f}_{\iota-1}} + \widehat{\alpha_1}\boldsymbol{v_i} \tag{11}$$

Optimal $\alpha_i (i \geq 2)$ can be found by using

$$\widehat{\alpha_\iota} = \arg\max\ NRIQA(\widehat{\boldsymbol{f}_\iota}) \tag{12}$$

C. Image Quality Assessment

Image quality assessment (IQA) is a technique to objectively evaluate an image's quality. When an IQA algorithm predicts an image's quality by comparing it with reference image, it is called Full-Reference IQA (FR-IQA). Examples of FR-IQA algorithms include MSE, PSNR, SSIM [25], FSIM [26], and GMSD [27]. The FR-IQA algorithms are useful for evaluating the effectiveness of an image restoration algorithm.

There is also a branch of IQA technique that evaluates an image's quality without any need of reference image: No-Reference IQA (NR-IQA). NR-IQA techniques usually compare the test image with their training database: a pair of image database and human subjective opinion score. However, recent development of NR-IQA bring forth a "completely blind" indexes that do not utilize human subjective opinion. In this paper, NR-IQA is utilized for determining optimal weight for restoration. NR-IQA and optimization technique chosen to optimize the weights highly affect the restoration quality of PCA-based method. Some example of NR-IQAs are BRISQUE [28], NIQE [23], IL-NIQE [29], and GM-LOG [30].

III. Proposed Method

A. Contribution ratio of principal components

Previously, we blindly use all the principal components by setting $J = M - 1$. Now, we conducted a small experiment to investigate the number of principal components contributing for restoration. The experiment was conducted using ten images that were degraded using Gaussian blur with varying strength, $\sigma_0 = 1, 1.25, 1.5, 1.75, 2$.Then, those images were restored using the cost function in Eq .(8) with a slight modification. PSNR, a FR-IQA is used to optimize the weights in place of NR-IQA. The parameters for ensemble generation were $M = 14, \sigma = 0.3, J = 13$.

In Fig. 2, the results of this small experiment are described. These results were from the average of all 10 images. The Y axis shows the contribution in terms of PSNR. The results on Fig. 2 show that in most cases only the first four principal components contribute to restoration process. In most cases, there is nearly no contribution from 5th and further principal components.

B. Sign determination and optimization algorithm

In a rare case, our previous algorithm fail to perform a restoration and instead further degrades the input image. This is caused by the combination of hill-climbing algorithm as an optimization method and the NR-IQA to optimize the restoration quality. Generally, plotting an NR-IQA index versus weight of principal component generates an oscillating plot. Occasionally, the starting point is the oscillating point, causing the hill-climbing to choose the wrong direction for the weight. To tackle this problem, we propose to use the dot product to determine the correct sign (direction of principal components) such that they do not degrade the image. The steps for determining the correct sign are:

1. Compute Δg by

$$\Delta g = g - \Psi \tag{13}$$

2. Compute the dot product of each principal components.

$$d_i = \Delta g \cdot v_i;\ i = 1{\sim}J \tag{14}$$

3. Determine the sign for each principal components

$$s_i = \begin{cases} 1 & d_i > 0 \\ -1 & \text{otherwise} \end{cases} \tag{15}$$

By applying this sign determination algorithm, the optimization technique need only to check in either positive or negative region. Because a standard hill-climbing algorithm cannot be "forced" to check into only a certain direction, a modification to the hill-climbing algorithm was needed.

We proposed to add an exhaustive search before applying hill-climbing search. The flowchart of our modified hill-

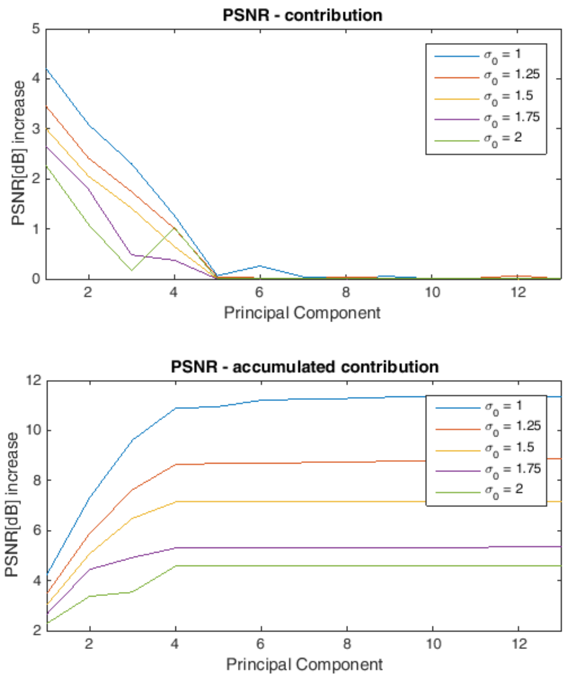


Fig. 2 Principal components' contribution in restoration process

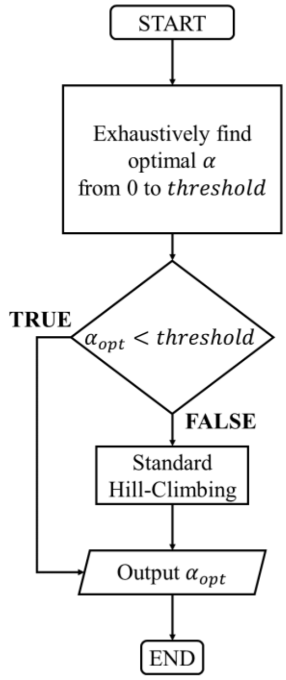


Fig. 3 Flowchart of modified optimization algorithm

climbing algorithm is shown on Fig. 3. The idea is basically to fail-safe our sign determination algorithm. If the optimal alpha is equal to threshold, then the surely the cost function can be further maximized. If the optimal alpha is less than threshold, than our sign determination algorithm may have failed; in that case we stop the optimization so no degradation occurs.

IV. Results and Discussions

The experiments were conducted to show the effectiveness of the proposed method. For benchmarking, we compared our proposed method with Levin et al.'s method [14], and Matlab's built-in function deconvblind, which is an implementation of maximum likelihood algorithm with assumption of additive Poisson noise. The restoration quality were compared in terms of PSNR and SSIM [25]. Other experiments were also conducted to investigate the noise robustness of our proposed method and the possibility for applying our method to other types of blur.

All restoration using our proposed and previous method use the following parameters for ensemble generation: M=14, σ=0.3 and NIQE [23] for optimizing the weights. We use 11x11 pixels as the kernel size for Levin et al's method. For simulated blur images, the Matlab's deconvblind, we supply the correct PSF as its initial guess.

All experimental data presented here were processed using a typical processor: Intel Core i3-2350M @ 2.30 GHz, MATLAB R2014b. Test images used for this experiment are obtained from USC-SIPI image database [31] and CVG-UGR image database [32].

A. Restoration quality of proposed method

The restoration quality of the proposed method is compared to the benchmarking algorithm. Three images were simulated blur images generated by convolving source image with Gaussian smoothing filter, $\sigma = 1.25, 1.5, 1.75$. The restoration results are shown in Table I. One more image was blurred with Gaussian smoothing filter, $\sigma = 1.3$ to show the rare case of failed restoration in previous method and our proposed method's success. This is shown on Fig. 4..

B. Noise Robustness

In this experiment, we also evaluate our algorithm's performance in the presence of noise. The noise is white Gaussian noise, with zero-mean and standard deviation of 10^{-3}. But, we found that the proposed method performs worse than previous method. Noise exists in the principal components and has higher amplitude than the restored signal making the weight optimization non-optimal. To tackle this problem, we proposed to apply Gaussian blur to the noisy image to remove noise from the image. In this experiment, we applied Gaussian smoothing filter $\sigma = 2$ to attenuate noise effect. The results of noisy image restoration are shown on Fig. 5.

C. Application to camera defocus blur

We also try to apply our proposed method to other types of blur, a camera defocus blur. The simulated blur images were degraded using 10 pixel-radius camera defocus blur and for optimizing the weights of proposed method, PSNR was used. Beside further Gaussian blur, we also tried further camera defocus blur for ensemble generation technique. Restoration results shown on Fig. 6 indicate the possibility of applying our

TABLE I. RESTORATION RESULT

		Degraded Image		Proposed Method		Levin et al.		Matlab's deconvblind	
Image	σ	PSNR [dB]	SSIM	PSNR [dB]	SSIM	PSNR [dB]	SSIM	PSNR [dB]	SSIM
Aircraft	1.25	29.32	0.9065	**35.55**	**0.9792**	21.85	0.7447	31.50	0.9419
	1.5	27.98	0.8776	**34.13**	**0.9676**	29.76	0.9098	30.29	0.9197
	1.75	27.01	0.8516	**32.93**	**0.9482**	29.02	0.8927	29.23	0.8954
Sailboat	1.25	27.74	0.8157	**36.25**	**0.9530**	20.25	0.5885	30.09	0.8802
	1.5	26.57	0.7781	**31.89**	**0.9053**	25.59	0.7795	28.88	0.8498
	1.75	25.70	0.7456	**27.89**	**0.8427**	25.43	0.7636	27.87	0.8187
Mandrill	1.25	22.65	0.6102	**25.35**	**0.8045**	19.36	0.3495	24.30	0.7724
	1.5	21.97	0.5398	**23.65**	**0.7022**	20.15	0.4178	23.20	0.6917
	1.75	21.52	0.4876	**23.11**	**0.6579**	20.17	0.4020	22.46	0.6209

(1) (2) (3) (4) (5) (6)

Fig. 4 Failed restoration case for previous method: (1) Original image, (2) degraded image, (3) restored with proposed method, (4) failed restored with previous method, (5) restored with [13], (6) restored with Matlab's deconvblind.

method for camera defocus blur, given correct ensemble generation parameters and weight optimization technique.

Previously, we found that optimal ensemble generation parameters for atmospheric-turbulence degraded images were M=14, σ=0.3. But, the results in Fig. 6 shows that it doesn't work for camera defocus images; and further camera defocus-ensemble instead gives higher restoration. We hypothesized that the optimal ensemble generation method has large M (more than 10) with small blurring PSF which has same shape with degrading PSF.

We also found that in camera defocus blur images, the contribution of principal components is different with Gaussian blur images. We perform an experiment using ten test images, degraded with camera defocus blur with varying strength: radius = 1px, 2px, 3px, … 10px. The ensemble generation uses camera defocus blur with 1px radius and we optimize the principal

(A1)
PSNR [dB] = 25.24, SSIM = 54.35%

(A2)
PSNR [dB] = 26.17, SSIM = 70.41%

(A3)
PSNR [dB] = 24.91, SSIM = 53%

(B1)
PSNR [dB] = 23.22, SSIM = 52.20%

(B2)
PSNR [dB] = 23.98, SSIM = 60.55%

(B3)
PSNR [dB] = 23.07, SSIM = 51.10%

Fig. 5 Restoration of noisy images: (A_) Boat, (B_) Stream and Bridge; (_1) Degraded and Noisy image, (_2) Proposed method with preprocessing, (_3) Proposed method without preprocessing.

components' weights using PSNR. In Fig. 7 we can see that for $J=13$, all principal components still contribute to some degree. We hypothesized that the contribution of principal components is related to blurring type.

V. Conclusions

This paper proposed an improvement to PCA-subspace restoration method for Gaussian blur by adding sign-determination algorithm and modifying the optimization algorithm. Experimental results show that our proposed improvement solved the no-restoration case without reducing the restoration quality for other cases. We also checked the noise robustness of our proposed algorithm and showed that our method worked after applying simple noise removal algorithm. In the last experiment, we checked the possibility of applying our proposed method to other types of blur. From this experiment we concluded two things. First, the method works best if the ensemble PSF is similar in shape with true degrading PSF. Second, the number of contributing principal components depend on the type of degradation. Our future works will be improving the noise robustness, adding artifact removal for camera misfocus blur, and to apply the proposed method for general blurring PSF.

Acknowledgment

B. Sumali would like to thank Japan-ASEAN Integration Fund (JAIF) for the MJIIT-ASEAN student incentive to pursue his study.

References

[1] R. C. Gonzalez, R. E. Woods, *Digital image processing*, 2nd ed. Prentice Hall, 2007

[2] D. Kundur and D. Hatzinakos, "Blind image deconvolution," in *IEEE Signal Processing Magazine*, vol.13, no.3, pp.43–64, 1996

[3] D. Kundur and D. Hatzinakos, "Blind image deconvolution revisited," in *IEEE Signal Processing Magazine*, vol. 13, no. 6, pp. 61–63, 1996.

[4] A. Kaur and V. Chopra, "A Comparative Study and Analysis of Image Restoration Techniques Using Different Images Formats," in *International Journal for Science and Emerging Technologies with Latest Trends*, vol. 2, no. 1, pp. 7–14, 2012.

[5] M. K. Rajput and R. K. Singh, "A Review ON IMAGE RESTORATION TECHNIQUES.", in *International Journal of Advanced Technology and Engineering Research*, vol. 5, no. 6, 2015

[6] R. G. Lane and R. H. T. Bates, "Automatic multidimensional deconvolution," in *Journal of Optical Society of America A*, vol. 4, no. 1, pp. 180–188, 1987.

[7] G. R. Ayers and J. C. Dainty, "Iterative blind deconvolution method and its applications," in *Optics letters*, vol. 13, no. 7, pp. 547–549, 1988.

[8] L. Lucy, "An iterative technique for the rectification of observed distribution," in *Astronomical Journal.*, vol. 79, no. 6, pp. 745–754, 1974.

Original Image: Goldhill

Original Image: Text

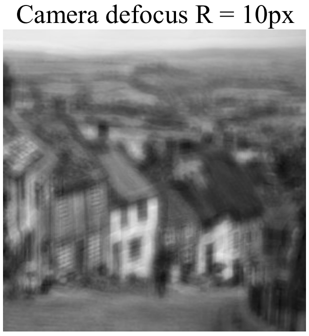
Blurred image

Matlab’s deconvblind
Initial PSF:
Camera defocus R = 10px

Blurred image

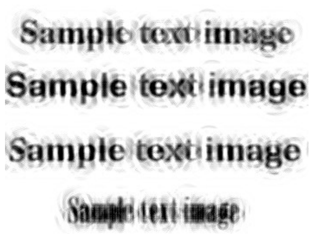
Matlab’s deconvblind
Initial PSF:
Camera defocus R = 10px

Proposed method:
M=14, J=13,
Gaussian Blur σ=0.3



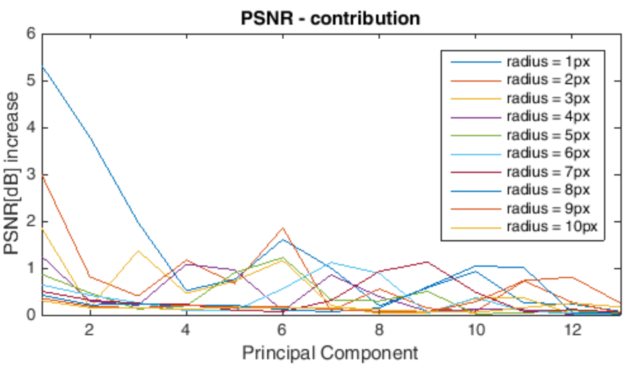
Proposed method:
M=14, J=13,
Camera defocus R = 1px

Proposed method:
M=14, J=13,
Gaussian Blur σ=0.3



Proposed method:
M=14, J=13,
Camera defocus R = 1px

Fig. 6 Restoration of camera defocus images



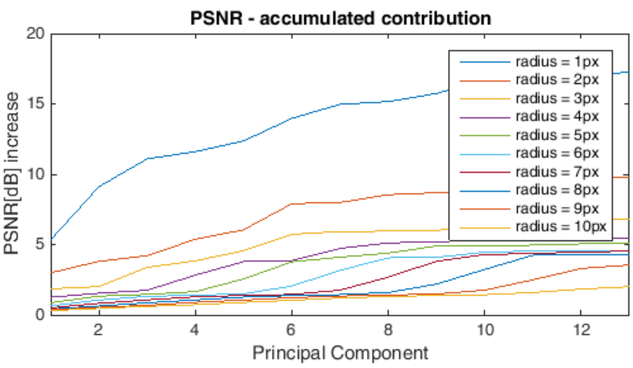


Fig. 7 Principal components’ contribution in restoration process

[9] W. H. Richardson, “Bayesian-based iterative method of image restoration,” in *Journal of Optical Society of America A*, vol. 62, no. 1, pp.55–59,1972.

[10] D. A. Fish, J. G. Walker, A. M. Brinicombe, and E. R. Pike, "Blind deconvolution by means of the Richardson–Lucy algorithm," in *Journal of Optical Society of America A*, vol. 12, no. 1, pp. 58–65, 1995

[11] R. Fergus, B. Singh, A. Hertzmann, S.T. Roweis, and W.T. Freeman, "Removing camera shake from a single photograph," in *The 33rd International Conference and Exhibition on Computer Graphics and Interactive Techniques (SIGGRAPH),* 2006.

[12] Ju-Ting Chang and Long-Wen Chang, "Image Decomposition and Patch Selection for Single Image Blind Deblurring", in *The Eighth International Workshop on Image Media Quality and its Applications (IMQA)*, Nagoya (Japan), 2016

[13] A. Levin, Y. Weiss, F. Durand, and W. T. Freeman, "Understanding and evaluating blind deconvolution algorithms," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1964–1971., 2009

[14] A. Levin, Y. Weiss, F. Durand, and W. T. Freeman, "Efficient marginal likelihood optimization in blind deconvolution," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* pp.2657–2664, 2011

[15] A. Levin, "Blind motion deblurring using image statistics," in *Advances in Neural Information Processing Systems (NIPS),* 2006.

[16] J. W. Miskin and D. J. C. MacKay, "Ensemble learning for blind image separation and deconvolution," in *Advances in Independent Component Analysis*, Springer, 2000.

[17] Dalong Li, S. Simske, and R. M. Mersereau, "Blind image deconvolution using constrained variance maximization," in *The Thirty-Eighth Asilomar Conference on Signals, Systems and Computers*, vol.2, pp.1762–1765 Vol.2, 2004

[18] Dalong Li, R. M. Mersereau, and S. Simske, "Atmospheric Turbulence-Degraded Image Restoration Using Principal Components Analysis," in *IEEE Geoscience and Remote Sensing Letters*, vol.4, no.3, pp.340–344, 2007

[19] R. Nakamura, Y. Mitsukura, and N. Hamada, "Iterative PCA approach for blind restoration of single blurred image," in *International Symposium on Intelligent Signal Processing and Communications Systems (ISPACS)*, pp.543–546, 2013

[20] R. Nakamura, Y. Mitsukura, and N. Hamada, "Blind restoration of single-channel image using iterative PCA," in *IEEE Conference on Systems, Process & Control (ICSPC)*, pp.84–87, 2013

[21] B. Sumali, N. Hamada, and Y. Mitsukura, "Blind Image Restoration Method by PCA-based Subspace Generation", in *International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS)*, pp. 204-209, Bali, Indonesia, 2015

[22] B. Sumali, N. Hamada, and Y. Mitsukura, "Blind Image Restoration using PCA and Image Quality Assessment", in *The Eighth International Workshop on Image Media Quality and its Applications (IMQA)*, Nagoya, Japan, 2016

[23] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a 'Completely Blind' Image Quality Analyzer," in *IEEE Signal Processing Letters*, vol.20, no.3, pp.209–212, 2013

[24] A. Jain, "Image Enhancement," in *Fundamentals of Digital Image Processing*, Prentice-Hall, 1989

[25] Zhou Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," in *IEEE Transactions on Image Processing*, vol.13, no.4, pp.600–612, 2004

[26] Lin Zhang, Lei Zhang, Xuanqin Mou, and David Zhang, "FSIM: a feature similarity index for image quality assessment", in *IEEE Transactions on Image Processing*, vol. 20, no. 8, pp. 2378–2386, 2011

[27] W. Xue, L. Zhang, X. Mou, and A. C. Bovik, "Gradient magnitude similarity deviation: a highly efficient perceptual image quality index," in *IEEE Transactions on Image Processing*, vol. 23, no. 2, pp. 684–695, 2014.

[28] A. Mittal, A. K. Moorthy and A. C. Bovik, " No-Reference Image Quality Assessment in the Spatial Domain ", in *IEEE Transactions on Image Processing*, 2012

[29] Lin Zhang, Lei Zhang, and A. C. Bovik, "A Feature-Enriched Completely Blind Image Quality Evaluator," in *IEEE Transactions on Image Processing*, vol.24, no.8, pp.2579–2591, 2015

[30] W. Xue, X. Mou, L. Zhang, A. C. Bovik, and X. Feng, "Blind image quality assessment using joint statistics of gradient magnitude and laplacian features," in *IEEE Transactions on Image Processing*, vol. 23, no. 11, pp. 4850–4862, 2014.

[31] University of Southern California. *USC-SIPI Image Database* [Online]. Available: http://sipi.usc.edu/database/

[32] University of Granada. *CVG–UGR Image Database Computer Vision Group* [Online]. Available: http://decsai.ugr.es/cvg/dbimagenes/

Predicting the Popularity of Online News using Gradient Boosting Machine

Md. Taufeeq Uddin[1], Mohammad Shahadat Hossain[3], Muhammed Jamshed Alam Patwary[2] and Tanveer Ahsan[2]

[1,2]Department of Computer Science and Engineering, International Islamic University Chittagong, Bangladesh
[3]Department of Computer Science and Engineering, University of Chittagong, Bangladesh
mohamed.taufeeq.uddin@gmail.com

***Abstract*—Popularity prediction of online news aims to predict the future popularity of news article prior to its publication estimating the number of shares, likes, and comments. Yet, popularity prediction is a challenging task due to various issues including difficulty to measure the quality of content and relevance of content to users; prediction difficulty of complex online interactions and information cascades; inaccessibility of context outside the web; local and geographic conditions; and social network properties. This paper focuses on popularity prediction of online news by predicting whether users share an article or not, and how many users share the news adopting before publication approach. This paper proposes the gradient boosting machine for popularity prediction using features that are known before publication of articles. The proposed model shows around 1.8% improvement over previously applied techniques on a benchmark dataset. This model also indicates that features extracted from articles keywords, publication day, and the data channel are highly influential for popularity prediction.**

***Keywords*—Text Mining, Social Media Contents, Popularity Prediction, Gradient Boosting Machine, Before Publication Approach**

I. INTRODUCTION

The consumption of online news accelerates day by day due to the widespread adoption of smartphones and the rise of social networks. Note that online news content comprises of numerous key properties, for instance, it is easily produced and small in size; and its lifespan is short and the cost is low. Such properties make news content more effective to be consumed on social sharing platforms. More interestingly, this type of content can capture the attention of a significant amount of Internet users within a short period of time. As a consequence, researchers focus on the analysis of online news content such as predicting the popularity of news articles, demonstrating the decay of interest over time to understand the world of online news since it has so many practical implications [1].

The prediction of the popularity of online news content has remarkable practical values in many fields. For example, by utilizing the advantages of popularity prediction, news organization [2] can gain a better understanding of different types of online news consumption of users. As a result, the news organization can deliver more relevant and engaging content in a proactive manner as well as the organization can allocate resources more wisely to develop stories over their life cycle. Furthermore, prediction of news content is also beneficial for trend forecasting, understanding the collective human behavior, advertisers to propose more profitable monetization techniques, and readers to filter the huge amount of information quickly and efficiently [1] [3].

The notion of popularity is often expressed by investigating the number of interactions in the web and social networks, for example, click-through rate, number of shares, likes, and retweets. Tatar et al. [4] demonstrated two types of popularity prediction techniques that are

1. **After publication technique:** more common technique, which uses features capturing the attention that one content receives after its publication. Higher prediction results are expected in after publication technique since utilization of information about the received attention makes the prediction task easier [1] [5] [6] [7].
2. **Before publication technique:** relatively challenging and effective technique. This technique uses only content metadata features that are known prior to the publication of contents instead of using features related to the attention that one content receives after contents release. Although the expected prediction accuracy is comparatively low in before publication method as we are using only metadata features rather than original news content [8], the prediction is more desirable as far as it fosters the possibility of decision making to customize the content before the release of content [9]. In this work, we model popularity prediction problem in before publication technique.

Although popularity prediction of web content has tremendous impacts in many areas, popularity prediction task still faces a bunch of major challenges [4] [9]. First, different factors make prediction difficult, for example, the quality of content or relevance of content to users can influence contents popularity. Second, the relationship between events in the real world and content itself are not only difficult to capture but also hard to further feed into the prediction engine. Third, prediction of complex social interactions and information cascades at the microscopic level are extremely challenging. Fourth, the prediction might also be difficult because of the inaccessible content like context outside the web, local and geographical conditions, and situations which influence the population. Last but not least, the prediction may also be hard based on the network properties e.g. the structure of the networks, and the interplay between different layers of the web.

Previously, researchers try to estimate the popularity by predicting whether or not someone share the news. However, this approach is less informative since we can only identify users share the news rather than how many users share the news. Hence, this paper proposes an extension to the previous popularity prediction models by predicting the number of shares of news using a novel ensemble learning algorithm, namely gradient boosting machine [10] in **before publication** setting. In this work, we use a heterogeneous set of metadata features that are known prior to the publication of the article to train the proposed gradient boosting machine, where the goal is not only to predict whether users share an article or not but also to predict how many users shares the article. Hence, gradient boosting machine is the ensemble of multiple weak learners or decision trees in which each successive decision trees are built from the prediction residual of the preceding decision trees to form a final highly accurate prediction model. The final prediction model guarantees that it performs much better than the individual performance of each decision tree. Note that the proposed gradient boosting machine comprises of several good qualities including high popularity prediction accuracy, competitive computational performance both in training and prediction steps, and capability to handle large training dataset.

In our experiments, we found that gradient boosting machine (**GBM**) discriminated popular news articles from unpopular articles with around **74.5%** accuracy on publicly available Mashable news article dataset [8] that was as much as **1.8%** improvement over the previous approach. In addition, gradient boosting machine measured that information regarding the summary of the articles keywords, data channel types, the earlier popularity of referenced articles, natural language processing features, and publication day of articles was crucial for predicting the popularity of online news. To the best of our knowledge, this is the first time that gradient boosting machine with linear and logistic regression loss functions is applied to the task of predicting the popularity of online news.

II. RELATED WORK

Over the last couple of years, researchers conducted several web mining and machine learning studies regarding web content analysis. For instance, Tatar et al. [4] analyzed different types of web content such as online videos, online news and social networking sites. They also reviewed different web content popularity prediction models including both classification and regression models. They further presented *good predictive features such as characteristics of content creators, textual features, and sentiment analysis, and revealed influential factors toward web content popularity*. Gao et al. [3] investigated the arrival process of retweets as well as user activity variation on the retweeting dynamics. They predicted the popularity by modeling the retweeting dynamics using extended reinforced Poisson process model with time mapping process. They, in addition, reduced the effect of user activity variation introducing the Weibo time notation as well as integrating a time mapping process into the proposed model. Castillo et al. [2] provided a qualitative and quantitative analysis of the life cycle of articles' stories demonstrating the interplay between site visitation patterns and social media reactions to articles. Furthermore, they modeled overall traffic of articles by observing social media reactions or attention profile such as *decreasing, steady, increasing,* and *rebounding*.

After publication method is highly popular in popularity prediction research. For instance, Tatar et al. [1] ranked news articles by predicting *user comments* using the linear model on a logarithmic scale and constant scaling model in after publication setting. They outlined that *popularity prediction methods are the good alternative for automatic online news ranking*. Lee et al. [5] inferred the likelihood of the popularity of online content for survival analysis applying Cox proportional hazard regression. They used a set of observable explanatory factors to model and to predict objective metric such as threads lifetime and the number of comments. Petrovic et. al. [11] predicted message propagation in twitter using passive-aggressive (PA) learning algorithm in time-sensitive approach. They extracted features related to the author and text of the tweets and various statistics of the tweet itself. Their findings suggested that *automatic retweets prediction performance of PA is as good as prediction performance of humans*. Szabo and Huberman [6] predicted the long-term dynamics of individual submissions e.g. number of views for Youtube videos and the number of votes for Digg stories from early measurements of access of users. They highlighted that Digg stories outdated shortly while Youtube videos were found popular for a long time after their initial submission to the portal.

However, there are only a few studies which followed the challenging *before publication approach* like this study. For example Bandari et al. [9] constructed features from the content of the news articles and its source of publication that were available prior to contents release. They considered four characteristics of the articles: *news source, the category of news, the subjectivity of the language,* and *named entities mentioned in the articles*. They reported that ranges of popularity on social media could possibly be predicted with *84%* accuracy using the bagging technique. Arapakis et al. [12] pointed out that *news popularity prediction at cold start is still an open challenge*. They predicted tweet counts and page views using features related to time, news source, genre, Wikipedia, web search and twitter. In their findings, they reported that *the imbalanced class distribution drove the prediction models to bias toward the unpopular articles that concluded the predictions not useful in realistic scenarios*. Fernandes et al. [8] proposed proactive intelligent decision support system for online news articles that predicted whether user shares articles or not analyzing features known before publication such as keywords, digital media content, and earlier popularity of news referenced in articles. They achieved 73% popularity prediction accuracy on Mashable news dataset via random forests algorithm adopting rolling windows evaluation strategy.

III. GBM FOR NEWS POPULARITY PREDICTION

The goal of this work is to predict whether a news article may share or not by users as well as the total count of shares in the realistic setting using GBM algorithm in both classification and regression settings. In training step of this approach, heterogeneous set of metadata features extracted from articles is fed to GBM to build the desired prediction model. The trained GBM model is then used in prediction step to carry out popularity prediction task. The rest of this section covers a brief description of the key idea, algorithmic construction, and regularization techniques of GBM for classification and regression.

A. Key Idea

GBM [13] is an ensemble learning algorithm that is the combination of gradient-based optimization and boosting. GBM produces strong prediction model by combining multiple weak prediction models in which weak models are created by sequentially applying to the incrementally changed dataset. Optimization based on the gradient in GBM utilize the gradient computations to minimize the cost function of a model with respect to training dataset while boosting additively gathers an ensemble of weak models to build the prediction model for popularity prediction challenge. In short, the main idea beneath GBM is to build a series of simple and probably inaccurate decision trees or weak models successively from the prediction residuals of the preceding decision trees and combine them to construct a final highly accurate prediction model.

B. Algorithmic Description

The technical description of the predictive GBM [10] [14] is given below:

- **Input:** GBM takes $(x_1, y_1), \ldots, (x_n, y_n)$ as training dataset in which $x_i \epsilon X$ are the set of features extracted from news content and y_i is the prediction label. GBM also takes a differentiable cost function $L(y, F(x))$ and the number of gradient boosting iteration M as input.
- **Algorithm:**
  I. Initialization of the model with a constant value

$$F_0(x) = argmin_\gamma \sum_{i=1}^{n} L(y_i, \gamma) \quad (1)$$

  II. For each gradient boosting iteration $m = 1, \ldots, M$

  1. For $i = 1, \ldots, n$, compute pseudo-residuals

$$r_{im} = -\left[\frac{\delta L(y_i, F(x_i))}{\delta F(x_i)}\right]_{F(x)=F_{m-1}(x)} \quad (2)$$

  2. For $j = 1, \ldots, J_m$, GBM fits a regression tree to the labels y_i providing terminal region R_{jm} . J is the size of the tree that controls the level of variable interaction in the model.
  3. For $j = 1, \ldots, J_m$, GBM then computes the multiplier γ_m

$$\gamma_{jm} = argmin_\gamma \sum_{x_i \epsilon R_{jm}} L(y_i, F_{m-1}(x_i) + \gamma) \quad (3)$$

  4. Finally, update the model

$$F_m(x) = F_{m-1}(x) + v \sum_{j=1}^{J_m} \gamma_{jm} I(x \epsilon R_{jm}) \quad (4)$$

  where v $(0 < v \leq 1)$ is the shrinkage constant that regulates the learning rate of GBM by reducing the size of the incremental steps.
- **Output:** $F_M(x)$

C. Regularization

Generalization capability of GBM is one of the crucial concern [15]. A number of parameters can contribute towards reducing the effects of overfitting by controlling learning rate and/or by introducing randomness into GBM. For instance, the smaller values of learning rate v such as $v \leq 0.1$ can generally ensure better generalization and performance on the test dataset. Overfitting can also be eliminated by fitting weak models on a subsample or constant fraction e.g. $0.5 \leq fraction \leq 0.8$ of the training dataset at random with no replacement. Furthermore, to reflect generalization, we can use a large number of trees or the boosting iteration M, and control J, by picking the value of J in between 4 and 8.

IV. MASHABLE NEWS DATASET

Online news popularity prediction (**Mashable news**) dataset [8] is publicly available at `http://archive.ics.uci.edu/ml/datasets/Online+News+Popularity`, which aims to predict the future popularity of news articles using information that are known before the release of news articles. Mashable news dataset consists of *58* heterogeneous features about the associated statistics of the original news articles released by Mashable (`www.mashable.com`) during a two years period from January *7, 2013* to January *7, 2015*. Fernandes et al. crawled news articles from Mashable website, and then they discarded special occasion and very recent articles e.g. < 3 weeks from the crawled articles.

More precisely, they extracted *47* features from HTML code and classified them into *4* different categories such as *number, ratio, bool*, and *nominal*. The unbounded numeric features like the number of words in the article were scaled by logarithmic transformation as well as they transformed the nominal features with the common *1-of-C* encoding. They also extracted the statistical summary of the number of shares of all Mashable links cited in articles that were known before the release of articles. Three types of keywords such as worst, average and best were captured by ranking all articles keyword average shares that were also known before release.

Additionally, they extracted a bunch of natural language processing features such as closeness to top latent Dirichlet allocation (LDA) [16] topics, title subjectivity, the rate of positive and negative words and title sentiment polarity by

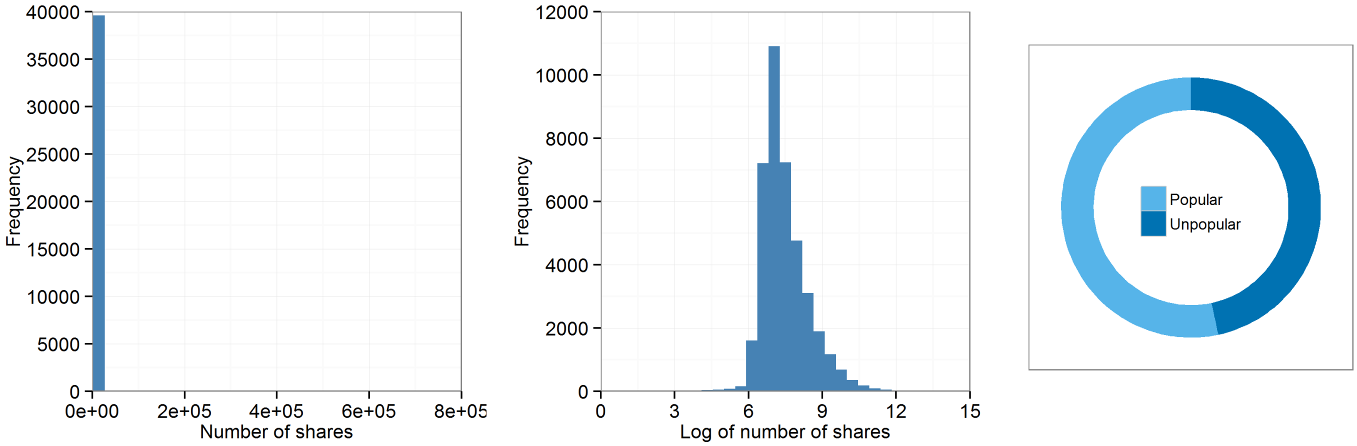


Fig. 1. From left to right, distribution of the popularity prediction label e.g. original number of shares, the logarithms of the number of shares, and proportion of the popular and unpopular article classes of Mashable news dataset

using LDA to compute relevant topics as well as to measure the closeness of current article to the previously computed topics. Sentiment polarity and subjectivity scores were also computed by applying the pattern web mining module [17].

V. EXPERIMENTS AND RESULTS

The proposed GBM prediction model was evaluated on Mashable news dataset using $5-fold$ cross-validation strategy. To benchmark the proposed GBM against previously applied algorithm, namely random forests, we also reproduced random forests (**RF**) from reference [8] adopting 5-fold cross validation. Each of the experiments was run for *20* times with different random seeds, and then, we averaged over *20* different experimental runs to achieve the final results. For predicting whether users share a news article or not, we modeled the popularity prediction problem as binary classification problem that is $Popular$ vs. $Unpopular$. In the case of binary classification, receiver operating characteristics (ROC) curve was produced to demonstrate the performance of the models. Note that ROC is created by plotting sensitivity against one minus specificity i.e. $1-specificity$ at numerous threshold values. The larger value of the performance metric area under the ROC curve (**AUC**) indicates the higher popularity prediction accuracy.

As mentioned earlier, the experimental set up was binary classification problem in which the goal was to classify whether an article was popular or unpopular by predicting whether users share that article or not. We defined the popularity of an article based on a decision threshold D. For instance, when an article is shared more than *1400* times e.g. $D \geq 1400$, we labeled the article as $Popular$; otherwise, we labeled the article as $Unpopular$ as suggested in [8]. We considered AUC as the primary evaluation metric as far as AUC is the most suitable metric as AUC is independent of the threshold value as well as AUC calculates discrimination power of the models very efficiently.

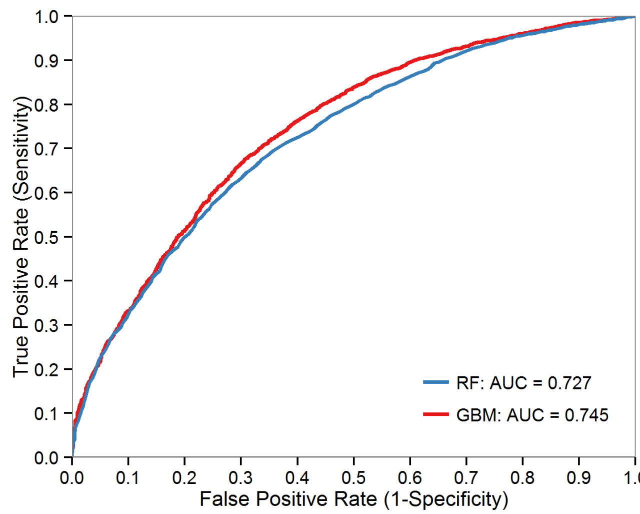


Fig. 2. Comparison of the AUC produced from RF and GBM

On the other hand, we modeled the popularity prediction problem as regression problem during the estimation of popularity via predicting the number of shares of the news article by users. Note that we took the logarithm of the original number of shares of news to use as the prediction label to train and test using GBM and RF regressors. During performance evaluation of the regression models, we evaluated the models by measuring the mean absolute percentage error between the predicted shares and logarithm of the original number of shares (**MAPE_LOG**), and between the exponents of predicted shares and the original number of shares (**MAPE**).

For training and validating the proposed GBM, we set *logistic regression*, and *linear regression* as objective functions during binary classification and regression, respectively, the value of the learning rate $v = 0.001$ to reflect better generalization, the size of the tree $J = 8$, and subsample

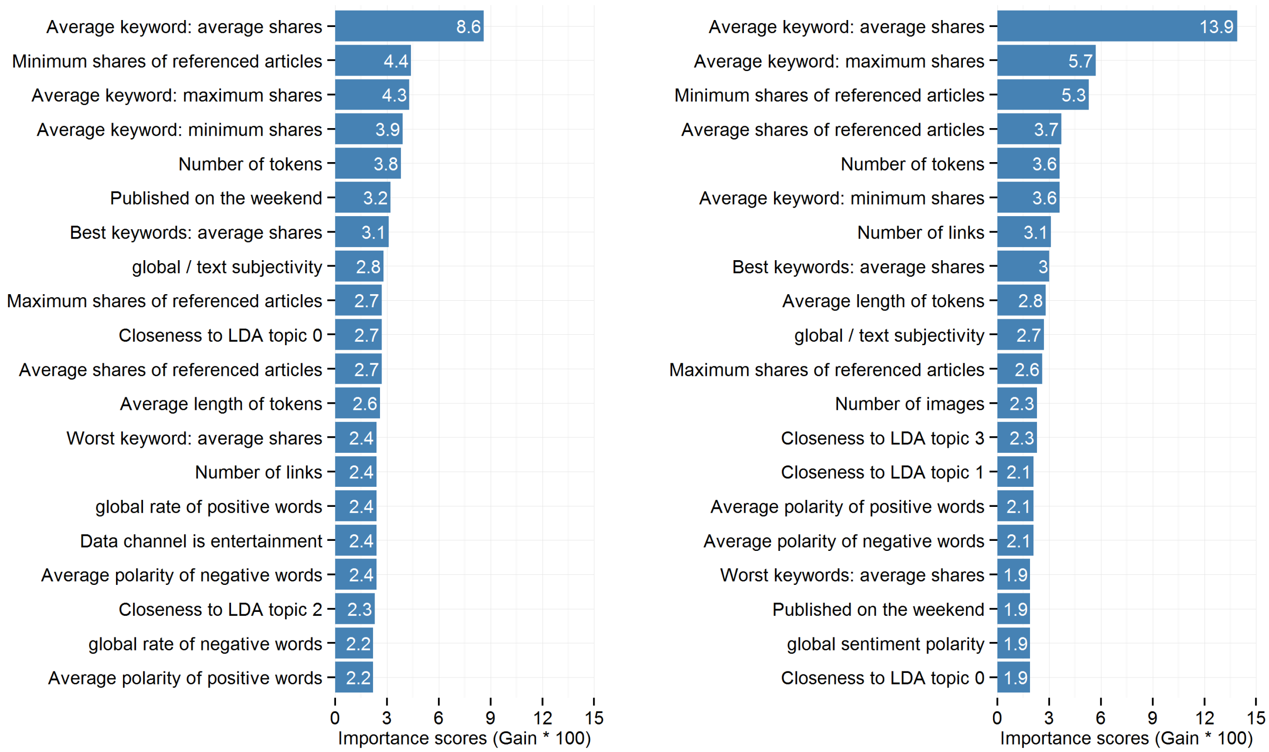


Fig. 3. Top 20 features based on the importance scores of features, measured via GBM, for predicting the, from left to right, popularity of online news, and number of shares of online news

TABLE I
PREDICTION BENCHMARK OF NUMBER OF SHARES OF NEWS ARTICLES. THIS TABLE DISPLAYS MAPE_LOG, MAPE, AND THEIR STANDARD DEVIATION (**SD**) RESULTED FROM RF AND GBM APPLYING ON MASHABLE NEWS DATASET

Model	MAPE_LOG	MAPE_LOG SD	MAPE	MAPE SD
RF	8.39 %	0.0083	73.41 %	0.0986
GBM	**8.11 %**	0.0073	**69.42 %**	0.0989

size $fraction = 0.8$. Finally, the number of boosting iteration M was selected using 5-fold validation approach. Figure 2 shows the performance of RF and GBM classifiers. Notice that the best AUC value obtained on Mashable news dataset was **74.5%** using GBM. It can be seen that GBM performed **1.8%** better than widely used machine learning algorithm RF which generated previous benchmark results on Mashable news dataset. Although the obtained result was around **74-75%** discrimination level which was far away from being perfect, performance was still interesting since we only fed metadata features for prediction task following before publication technique. As we obtain this prediction results using some meta features or statistical features before publication of the news, we can, therefore, feed this results for further modification of the articles.

Table I shows the performance of RF and GBM during the prediction of the count of shares of articles. From Table I, it can be indicated that GBM was better than RF since MAPE_LOG and MAPE generated from GBM were 8.11% and 69.42%, respectively, that were relatively smaller than the MAPE_LOG and MAPE generated from RF.

GBM uses *gain* [18] to estimate the contribution of each feature to the prediction model. GBM takes each gain of each feature of each tree and computes mean per feature to provide a vision of the entire prediction model. We measured the relative importance scores of features using GBM for both predicting popularity and number of shares of news for which we trained GBM using *35679* news articles. Figure 3 highlights the relative importance scores or $Gain * 100$ of the top *20* features; x-axis represents the relative feature importance scores and y-axis represents the *20* features from Mashable news dataset. In can be observed from Figure 3 that features related to the summary statistics of contents keywords, the summary statistics the number of the shares of the referenced articles, tokens, links, data channel types whether it is entertainment or technology channels, and publication day had strong significance for predicting popularity and number of shares of online news. Moreover, natural language processing features such as closeness of target article to different LDA topics e.g. 0, 1, 2, rate of positive and negative words, text subjectivity, and global sentiment polarity encapsulated high discrimination power towards popularity prediction.

VI. CONCLUSION AND FUTURE WORK

This paper introduces and implements gradient boosting machine to tackle the challenge of classifying popular news articles from unpopular articles by measuring the count of shares in before publication approach. Our findings suggest that gradient boosting machine is able to predict popularity with a decent prediction rate using only statistical features associated with original news articles without using the original content of news articles or after publication attention. GBM also outlined discriminative and useful metadata features such as the statistical summary of keywords, the earlier popularity of articles referenced in articles, natural language processing features, and publication time. Future work will include, first, the exploration of more advanced features regarding content like trend analysis. Second, the evaluation of the prediction model on more complex and more unbalanced popularity prediction datasets. Third, the comparison of the model with many other state-of-the-art techniques.

References

[1] A. Tatar, P. Antoniadis, M. Amorim, and S. Fdida, "From popularity prediction to ranking online news," *Social Network Analysis and Mining*, vol. 4, no. 1, 2014. [Online]. Available: http://dx.doi.org/10.1007/s13278-014-0174-8

[2] C. Castillo, M. El-Haddad, J. Pfeffer, and M. Stempeck, "Characterizing the life cycle of online news stories using social media reactions," in *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing*, ser. CSCW '14. New York, NY, USA: ACM, 2014, pp. 211–223. [Online]. Available: http://doi.acm.org/10.1145/2531602.2531623

[3] S. Gao, J. Ma, and Z. Chen, "Modeling and predicting retweeting dynamics on microblogging platforms," in *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining*, ser. WSDM '15. New York, NY, USA: ACM, 2015, pp. 107–116. [Online]. Available: http://doi.acm.org/10.1145/2684822.2685303

[4] A. Tatar, M. de Amorim, S. Fdida, and P. Antoniadis, "A survey on predicting the popularity of web content," *Journal of Internet Services and Applications*, vol. 5, no. 1, 2014. [Online]. Available: http://dx.doi.org/10.1186/s13174-014-0008-y

[5] J. G. Lee, S. Moon, and K. Salamatian, "Modeling and predicting the popularity of online contents with cox proportional hazard regression model," *Neurocomputing*, vol. 76, no. 1, pp. 134–145, 2012.

[6] G. Szabo and B. A. Huberman, "Predicting the popularity of online content," *Communications of the ACM*, vol. 53, no. 8, pp. 80–88, 2010.

[7] M. Ahmed, S. Spagna, F. Huici, and S. Niccolini, "A peek into the future: Predicting the evolution of popularity in user generated content," in *Proceedings of the Sixth ACM International Conference on Web Search and Data Mining*, ser. WSDM '13. New York, NY, USA: ACM, 2013, pp. 607–616.

[8] K. Fernandes, P. Vinagre, and P. Cortez, *Progress in Artificial Intelligence: 17th Portuguese Conference on Artificial Intelligence, EPIA 2015, Coimbra, Portugal, September 8-11, 2015. Proceedings*. Cham: Springer International Publishing, 2015, ch. A Proactive Intelligent Decision Support System for Predicting the Popularity of Online News, pp. 535–546.

[9] R. Bandari, S. Asur, and B. A. Huberman, "The pulse of news in social media: Forecasting popularity," *CoRR*, vol. abs/1202.0332, 2012. [Online]. Available: http://dblp.uni-trier.de/db/journals/corr/corr1202.html#abs-1202-0332

[10] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *Annals of Statistics*, vol. 29, pp. 1189–1232, 2000.

[11] S. Petrovic, M. Osborne, and V. Lavrenko, "Rt to win! predicting message propagation in twitter." in *ICWSM*, L. A. Adamic, R. A. Baeza-Yates, and S. Counts, Eds. The AAAI Press, 2011. [Online]. Available: http://dblp.uni-trier.de/db/conf/icwsm/icwsm2011.html#PetrovicOL11

[12] I. Arapakis, B. Cambazoglu, and M. Lalmas, "On the feasibility of predicting news popularity at cold start," in *Social Informatics*, ser. Lecture Notes in Computer Science, L. Aiello and D. McFarland, Eds. Springer International Publishing, 2014, vol. 8851, pp. 290–299. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-13734-6_21

[13] M. M. Cliff Click, Jessica Lanford and V. Parmar, "Gradient boosted models with h2o's r package," February 2015: Second Edition, published by H2O.ai, Inc. 2307 Leghorn Street Mountain View, CA 94043, Inc. Available: https://leanpub.com/gbm/read. [Online]. Available: https://leanpub.com/gbm/read

[14] J. Friedman, T. Hastie, and R. Tibshirani, *The elements of statistical learning*. Springer series in statistics Springer, Berlin, 2001, vol. 1.

[15] "Gradient boosting," wikipedia, Wikimedia Foundation, Inc. , Web. Last Accessed: 26 August 2015. [Online]. Available: https://en.wikipedia.org/wiki/gradient_boosting

[16] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *the Journal of machine Learning research*, vol. 3, pp. 993–1022, 2003.

[17] T. De Smedt, L. Nijs, and W. Daelemans, "Creative web services with pattern," in *Proceedings of the Fifth International Conference on Computational Creativity*. Citeseer, 2014.

[18] tianqi chen, kailong chen, and tong he, "Xgboost," 2015, http://mloss.org/software/view/543/.

Design and Performance Analysis of CNFET Oscillator and Comparison with its CMOS Implementation

A.Q.M Shaiq [1], Md.Samiul Alam [2*] ,Md. Safwan Bin Alam[1],
[1]Department of EEE, American International University – Bangladesh (AIUB)
[2]School of Science, Engineering and Technology –EDU- Chittagong Bangladesh
*E-mail: samiulz.alam@gmail.com

Abstract **– The main objective of this paper is to design an oscillator using Carbon Nanotube Field Effect Transistor (CNFET) and evaluate its performance against the same oscillator implemented using CMOS. CNFET devices are widely considered faster much faster than their CMOS counterpart. Thus, the rise time, fall time and frequency of the CNFET oscillator was observed and compared with that of the CMOS oscillator under the same conditions. The results showed that the CNFET oscillator had much higher frequency and therefore a higher clock than then of the CMOS oscillator. All simulations are done using the HSpice model of CNFET.**

Keywords **– Carbon Nanotube Field Effect Transistor; Oscillator; HSpice; Transient Response;**

I. INTRODUCTION

Oscillators are one of the most used devices in electronics. From radios and TV transmitters to advanced microprocessors and video games, they all rely on oscillators. Most traditional oscillators are made of Silicon based Metal Oxide Semiconductor (MOS) technology. However, according to Moore's law, Complementary Metal Oxide Semiconductor (CMOS) technology has continued to scale down and are now reaching nano range dimensions such as 22nm [1]. In such dimensions these devices face several non-ideal effects such as low trans-conductance, gate oxide leakage, source to drain tunnelling, lesser ON-current, mobility degradation, increased delay etc. resulting to decreased gate control, leakage currents that are exponentially rising, high power density etc. Thus, eventually creating a large impact on reliability and cost of the devices, and making further scaling difficult [2].

Thus, nano-scaled-alternatives to the silicon transistor are being researched e.g. the carbon nanotube (CNT) based transistors, i.e. Carbon Nanotube Field Effect Transistors (CNFETs) [3]. These new materials and devices are expected to replace silicon and become commonly used transistors from the year 2015 (as the ITR predicts). The CNTs are referred to as the building blocks for silicon circuits [4]-[7]. CNFET devices have many advantages like a unique 1-D band structure that suppresses back scattering and makes near-ballistic operation possible. The quasi-1-D structure also provides better electrostatic control over the channel region than 3-D device and 2-D device structures [8]–[12].

Among the types of CNFETs the MOSFET-like CNFET (shown in Fig.1) has better device performance and fabrication feasibility compared to the SB-controlled [13]-[14]. Thus it was chosen for this paper, where the high speed of CNFETs was used to design a oscillator with ultra-fast performance compared to its CMOS implementation.

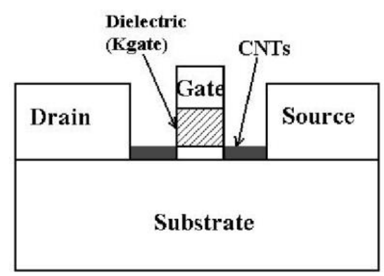


Fig. 1 Representation of MOSFET-like CNFET

In this paper, Section II is a brief introduction about the CNFET describing the structure and characteristics of the device. Section III underlies the simulation of the equivalent model of the CNFET, and expressions for the current sources and gate capacitances are given. Section IV presents the proposed oscillator design. In Section V the performance parameters of the CNFET oscillator is compared with that of the CMOS implementation. Finally in Section VI the paper has been concluded.

II. CARBON NANOTUBE FIELD EFFECT TRANSISTOR

The CNFETs overcomes most of the fundamental limitations for traditional silicon MOSFETs. With ultra-long (~1μm) mean free path (MFP) for elastic scattering, a ballistic or near-ballistic transport can be obtained with an intrinsic CNT under low voltage bias to achieve a better device performance. The ballistic transport operation and low OFF current make the CNTFET a suitable device for high performance and low power electronic oscillators. Moreover, the MOSFET-like CNTFET is likely to be scaled down to below 9 nm channel length, thus providing a substantial performance and power improvement compared to the MOSFET model (with minimum channel length of 32 nm).

MOSFET-like CNFETs have similar properties and operating principles as the Si MOSFETs. They have four terminals (drain, gate, source and substrate). Heavily doped CNT segments are placed between the drain, gate and source, and form the conducting channel and crates a low series resistance during the ON-state [18]. A dielectric film covers part of the undoped semiconducting CNT and the metal gate covers the dielectric. The gate controls the amount of charge in the channel via a vertical electric field which induces either electrons or holes in the nanotube. Meanwhile a horizontal electric field between the contacts provides the force that drives the charges from one contact to the other, resulting in an electric current. The threshold voltage of the intrinsic CNT channel is [17]:

III. Simulation Model Of CNFET

Fig.2 shows the equivalent circuit model implemented in HSPICE, which was proposed in [18].

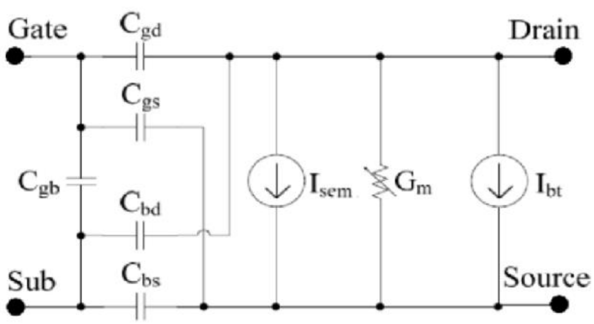


Fig. 2 Equivalent circuit model of CNFET

From the equivalent model, the three current sources are the thermionic current contributed by the semiconducting sub-bands (I_{sem}) with the classical band theory, the current contributed by the metallic sub-bands (I_m) that includes both electron and hole current, and the leakage current (I_{bt}) caused by the band to band tunnelling mechanism through the semiconducting sub-bands.

The macroscopic thermal conductivity is defined from Fourier's law of heat flow under nonuniform temperature. The steady state heat flow Jq is obtained by keeping the system and reservoirs in contact.

IV. DESIGN AND CIRCUIT DIAGRAM OF THE PROPOSED OSCILLATOR

An oscillator is a simple electronic circuit which produces a repetitive signal of waveforms often as sin wave or pulses. It has a feedback loop which feeds the output back to the input though selective filters to provide a positive feedback.

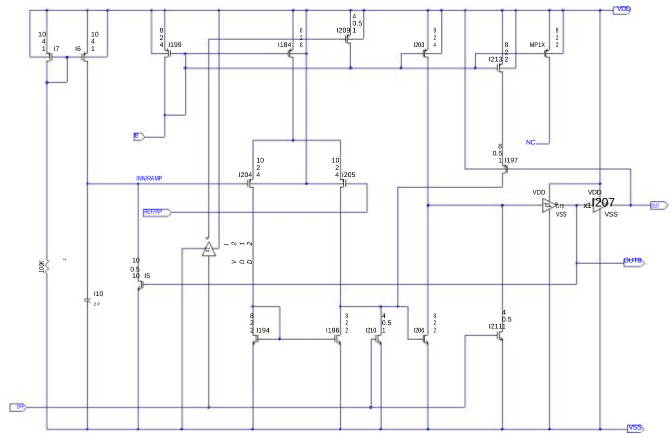


Fig. 3 Circuit design of the Oscillator used

In this design shown in Fig.3, a current mirror is formed by using Vdd, the current is copied to the inn pin of a comparator circuit, a DC reference voltage is applied to the other pin of the comparator. The comparator compares the voltage and generates an output (high or low) depending of the input in the differential pair. The output voltage is feedback to the inn pin of the comparator by using an NMOS/N-CNFET to copy the output current as a result a Ramp signal is formed at one of the inputs of the comparator which is again compared against the reference voltage and thus an oscillating circuit is formed.

V. SIMULATIONS, RESULTS, ANALYSIS AND DISCUSSION

Both the CNFET and CMOS oscillators were simulated on the Hspice platform using the Hspice CNFET model and the 0.5µm CMOS model. The observed parameters are shown in the following figures:

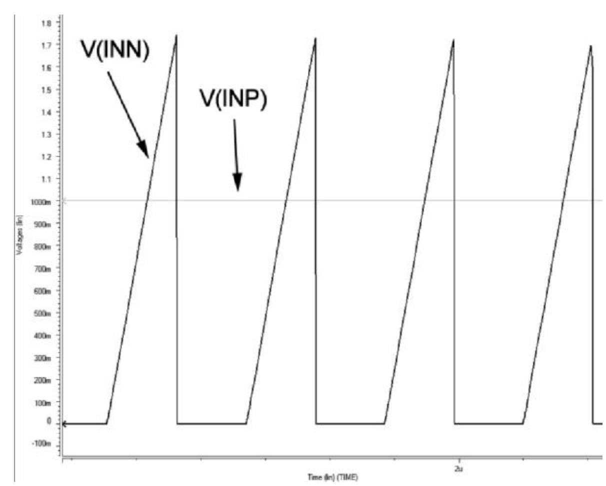


Fig. 4 Voltage vs. Time graph for CMOS Oscillator showing inputs in the V(inn) and V(inp) pins of the comparator

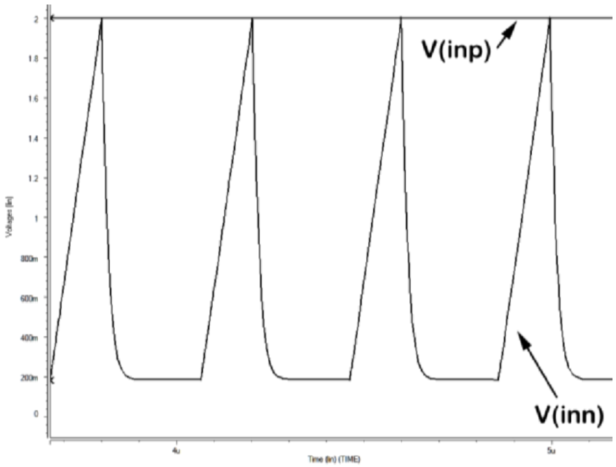


Fig. 5 Voltage vs. Time graph for CNFET Oscillator showing inputs in the V(inn) and V(inp) pins of the comparator

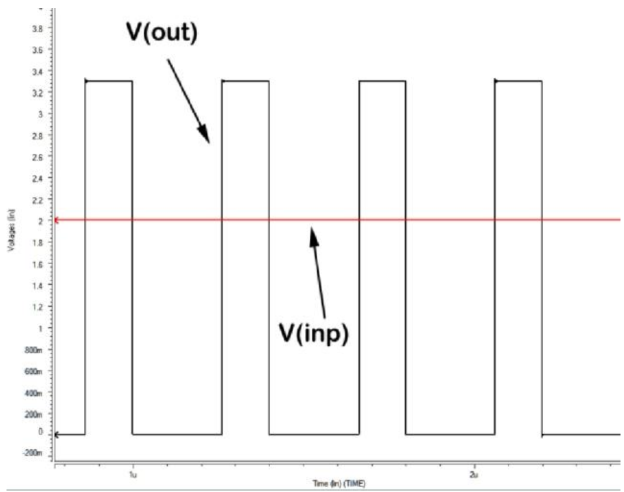


Fig. 6 Voltage vs. Time graph for CNFET Oscillator showing the output oscillating waveform.

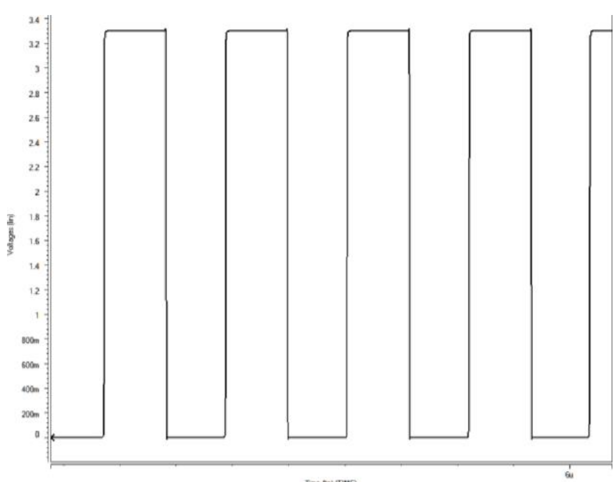


Fig. 7 Voltage vs. Time graph for CMOS Oscillator showing the oscillating output

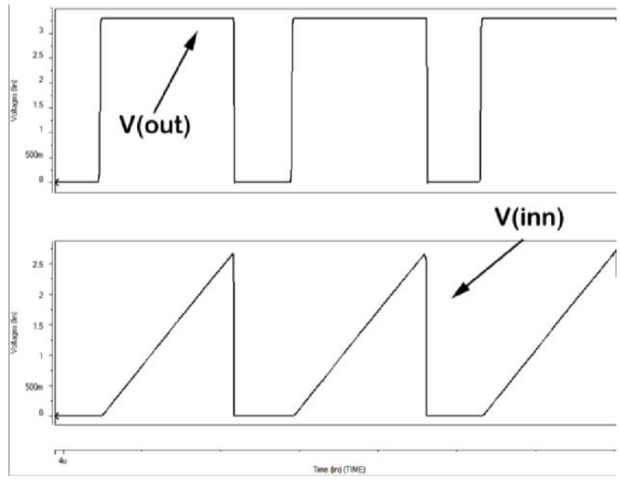


Fig. 8 Voltage vs. Time graph for CMOS Oscillator showing variations in Vinn and Vout

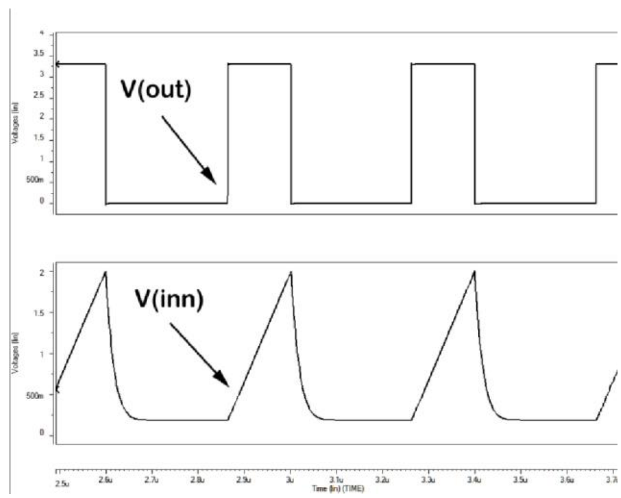


Fig. 9 Voltage vs. Time graph for CNFET Oscillator showing variations in Vinn and Vout

VI. CONCLUSION

This paper adequately designs and simulates both type of oscillator using CNFET and CMOS. From the simulation and matching results of waveform at various parts of the circuit we can conclude that it is possible to design similar circuits with CNFETs as CMOS without much alteration, it is because of the similarities between MOSFET-like CNTFETs and MOSFETs in terms of inherent characteristics and operation. The HSPICE simulation shows significant improvements in terms of speed and energy efficiency are achievable in different test conditions by utilizing the proposed design circuit in CNFET.

REFERENCES

[1] A Semiconductor Industry Association International Technology Roadmap for Semiconductors—2008 Update:Overview and Summaries,2008.http://www.itrs.net/Links/2008ITRS/Home2008.h`2] H.-S. P. Wong, "Beyond the conventional transistor", Solid-tate Electronics, vol. 49, issue 5, pp. 755-762, 2004.

[3] S.A.Tawfik, Z. Liu, V.Kursun,"Independent-gate and tied-gate FinFET SRAM Circuits: Design guidelines for reduced area and enhanced stability," Microelectronics, 2007. ICM 2007.Internatonal Conference on, pp. 171 - 174, Dec. 2007.

[4] Tans, S.A. Verschaerea and C. Dekker, "Room temperature transistor based on a SWCN," Science, 397: 49-52, 1998.

[5] Martel, R., T. Schmidt, H. Shea and T. Avouris, "S-and M-WCNFET Si transistors," Applied physics Letter, 73: 24 47-49, 1998.

[6] S.Heinze, J. Tersoff, R. Martel, V. Derycke, J.Appenzeller, and P.Avouris, "Carbon nanotubes as Schottky barrier transistors,"Phys. Rev.Lett., vol. 89, no. 10, pp. 106 801-1–106 801-4, 2002.

[7] Fabien Pr´egaldinyet.al., "Design Oriented Compact Models for CNFETs," IEEE Trans. Elec.dev., 2006.

[8] A. Rahman, J. Guo, S. Datta, M.S. Lundstrom, "Theory of ballistic nanotransistors," *IEEE Trans. Electron Devices*, vol. 50, no. 10, pp. 1853 - 1864, Sept. 2003

[9] A. Akturk, G. Pennington, N. Goldsman, "A. Wickenden, Electron Transport and Velocity Oscillations in a Carbon Nanotube,"*IEEE Trans. Nanotechnol.*, Volume 6, Issue 4, pp469 – 474, July 2007.

[10] H. Hashempour, F. Lombardi, "Device Model for Ballistic CNFETs Using the First Conducting Band," *IEEE Des. Test. Comput.*, Vol. 25, Issue 2, pp 178-186, March-April 2008.

[11] Y. Lin, J. Appenzeller, J. Knoch, P. Avouris, "High-performance carbon nanotube field-effect transistor with tunable polarities,"*IEEE Trans. Nanotechnol.*,Vol 4, Issue 5, pp 481 - 489, Sept. 2005.

[12] Peng, N., H. Li and Q. Zhang, 2009. "Nano Scale contacts between carbon Nanotubes and metallic Pads, "Acsnano, 10: 1016-25. 11.

[13] A. Bahari, G. Arasteh and S. Taghizadeh, ‖Nanometer-Scaled Contacts in CNFETs,‖ World Applied Sciences Journal 7 (Supplement 1): 18-21, 2009

[14] J. Deng and H.-S P. Wong, —A Compact SPICE Model for Carbon Nanotube Field Effect Transistors Including Non-Idealities and Its Application — Part I: Model of the Intrinsic Channel Region,‖ Submitted to *IEEE Transactions on Electron Devices*, 2007.

[15] S. Iijima,, "Helical microtubules of graphitic carbon", Nature, 354, 56-8, 1991.

[16] Bhushan, B., Ed. *Springer handbook of nanotechnology New York:*Spinger-Verlag Berlin Heidelberg, 2003.

[17] Stanford University CNFET Model Website.(2008). [Online]. Available: http://nano.stanford.edu/model.php?id=23

[18] J. Appenzeller, *"Carbon nanotubes for high-performance electronics Progress and prospect,"* Proc. IEEE, vol. 96, no. 2, pp. 201–211, Feb.2008.

An Environmental Sensing Experiment Using IEEE 802.15.4 Radio Hop of the WSN TelosB Motes

Yusnaidi Md Yusof[1], A.K.M. Muzahidul Islam[2], Sabariah Baharun[3], Mohd Ghazali Hamza[4], Kamarul Ariffin Abdul Basit[5], Noor Azurati Ahmad[6]

[1,2,3]MJIIT, [1,6]AIS, [4]Razak School of Engineering
Universiti Teknologi Malaysia (UTM)
Kuala Lumpur, Malaysia

{yusnaidi, muzahidul, sabariahb}@utm.my
ehkmohamed2@live.utm.my

[5]Faculty of Computer & Mathematical Sciences
Universiti Teknologi MARA (UiTM)
Shah Alam, Malaysia

k_rol@fskm.uitm.edu.my

***Abstract*—Wireless Sensor Networks (WSN) is a key technology to the pervasive monitoring and wireless surveillance. As part of the Internet of Things (IoT) and the Wireless Embedded Internet, WSN has shown a significant contribution to the remote control and monitoring of the environmental condition such as temperature, humidity, light, and acceleration detection. A sensor node or mote is a tiny-scaled electronic device with very limited power, memory, and bandwidth capable to sense, process, and transmit environmental data to the remote observer. This paper demonstrates the experience explored and the observation of temperature sensing using the IEEE 802.15.4 information hopping from a sensing mote to the base station. The environmental readings are furthered reprocessed and displayed on the base station screen for analysis. The study outcome and the experience gathered from the experiment are important to help the researchers to incorporate the information display, with the use of Wireless Embedded Internet's IPv6 and 6LoWPAN standard, to be globally accessible anywhere and at anytime.**

Keywords—WSNs; WSN data gathering; WSN data sensing; wireless information processing

I. Introduction

A Wireless Sensor Network (WS) is a group of specialized autonomous sensor nodes with a communication infrastructure to monitor environmental conditions such as temperature, humidity, light, acceleration, vibration, etc. The sensor nodes work cooperatively and are usually kept at the place or area in which data is expected to be sensed for monitoring and tracking purposes. The collected data is then passed to nearby neighbor sensor nodes in the same network until it arrived at the base station. The base station will have the opportunity to view all collected data and later converted them into understandable information so the observer would use it to make monitoring or any other decisions [1], [2].

A common architecture of WSN consists of three components: (1) sensor nodes, (2) sink node or base station, (3) personal computer [3]. A sensor node (mote) is a node in a sensor network which has a capability to perform processing, gathering sensory information as well as communicating with other connected nodes in the network. Among the main components of a sensor node are microcontroller, transceiver, memory, power source, and sensors [1]. The sensors in a sensor node is used to capture data from their environment. Many of the latest sensor node products implant the sensor chipset inside the sensor board itself. The sensor componets measure response to a change in a physical condition such as temperature or pressure. The captured data is communicated to a microcontroller component in the same sensor node for data processing. The microcontroller is able to process the data and controls the functionality of other components in the sensor node. The microcontroller are hardwired with the transceiver in the sensor board first to process and second to send the captured data to other neighboring sensor nodes [4], [5].

The sink node or also known as base station is a node responsibles to gather all sensed data coming from other nodes in the communication medium; and relay them to the observer whose can be located either local to the base station or at other remote area in which the connection has to go through the Internet [6], [7]. In a very naked understanding, the base station merely the same sensor node but with different functionality and is attached to the monitoring computer [2].

The personal computer to which the base station is attached to is the machine that used by the observer to monitor the sensing information coming from the base station. It can be located either locally to the sensing area or remotely at other location far from the sensing area where accessing this computer need an Internet connection. In a tsunami monitoring application for example, a bunch of thousand of thousands of sensor nodes are deployed in the deep ocean where the tsunami is predicted to originate. As the pressure level information are gathered about the the tsunami, these informations are wirelessly hopped-out from one node to another towards the base station. As the information reached the observer from the base station, an alarm action can be made to deal with the tsunami fenomena or threat. Other life-threatened events such as flight crash, habitat tracking, and health monitoring, to name a few, are all depends on how fast and how accurate the information can get to the human attention so the action can be made impromptu before the incidents happen. The above requirements are bigger required especially when human dealing with realtime data that need to be timely monitored [8].

As we can see the importance of the data capturing and monitoring, therefore the sensitivity of the captured data is at a paramount need. False translation and interpretation of data or unacceptable delay of data delivery at the observer site could threat human life in a bigger danger. The data gathered also need to pass to the observer at a faster rate (expecially when passing realtime data through the web) so the decision on the events can be made quicker.

This paper describes our works to demonstrate environmental data sensing and collection from autonomous sensor nodes and being readable with meaningful information on the observer screen.

The following sections describe the study in details. Section II defines standard and basic radio concept of data transmission in wireless sensor node. In Section III, we demonstrate the conducted experiment with results analysis. Finally, we envisage future works and some concluding remarks in Section IV.

II. IEEE 802.15.4 Radio Communication

IEEE 802.15.4 is the IEEE radio protocol standard for the communication of the ultra-low rate, ultra-low power consumption, and ultra-low bandwidth, and low cost devices and suitable for communication of sensor nodes in the WSN. It specifies the Medium Access Control (MAC) sublayer and physical layer for Low-Rate Wireless Private Area Networks (LR-WPAN) [9].

The features of low-cost, low-power of the IEEE 802.15.4 are intended to enable the promotion and deployment of WSN with the capability to live years on battery power together with mass deployment because of the very low cost of sensing devices [10]. The standard is optimized for low data throughput such as 250 kbps as in the low-bandwidth TelosB mote.

III. Implementation

In this section we present the environmental data sensing experiment using the IEEE 802.15.4 radio. The experiment is conducted in an office room of an eleven floor building.

A. Experiment Background

The experiment is conducted in an office room equipped with a standalone unit of air conditioner. The aim of the experiment is to sense the temperature of the room using the TelosB motes and monitor the reading at the screen of a computer where the base station is attached to.

The base station is installed with a TinyOS application which let the mote to relay any data coming from nearby motes to the computer either using serial or radio transmission. The application is called Basetation and is made available by the TinyOS [11].

We develop an application using the TinyOS nesC codes to be used on the sensing mote. The application has several features to enable and aid the mote in making a sensing task such as sensing the temperature of the room, hopping the sensed data to the base station, led indicator to signal the environmental data are being collected, and also a simple analysis program to display the sensed data on the base station computer screen.

Apart from the above mentioned characteristics, we also make used of the internal TinyOS tools and programs to display the collected data on the computer screen, namely ‘Listen’ and ‘PrintfClient’. The Listen java tool is used to display the raw sensing data together with the serial packet header bytes, while the PrintfClient java tool is used to display character strings of the temperature program output we made on the sensing mote.

In brief, our main aim of the experiment is to sense the temperature readings, hopping the data using the IEEE 802.15.4 ZigBee radio from the sensing mote to the base station, and viewing the readings on the computer screen to where the base station is attached. The communication between the sensing mote to the base station is through wireless radio while the communication between the base station and the observing computer is using serial cable transmission.

B. Hardware Setup

In the study, we use TelosB motes to gather temperature data and display the readings on the computer screen. TelosB mote is a ‘tiny form factor computer’ consists of microprocessor for processing, memory for data storage, and a number of sensors for data sensing. It has a dimension of 8cm x 3cm, capable of transmitting data at 250 kbps, has an integrated on-board antenna, and is IEEE 802.15.4 compliant. TelosB mote using the 8MHz TI MSP430 Microcontroller with a 48K bytes flash memory and 10K bytes RAM. It is an open source platform with a 16K bytes configurable EEPROM that allow programming and data collection via a built-in USB [12], [13], [14].

TelosB mote has an integrated Sensirion SHT11 temperature and humidity sensors [15], and also the Hamamatsu S1087 visible light and visible to IR sensors. The Sensirion SHT11 temperature sensor has a documented reading range of -40°C to 123.8°C with an accuracy of ±0.5°C, and sensing resolution of 0.01°C. Apart from that, the computer we are using is a 1.6 GHz Intel Core i5, 8 GB 1600 MHz RAM, and running OS X EI Capitan version 10.11.3.

We conduct the experiment by considering the following motes setup: fix and stationary mote, use IEEE 802.15.4 channel 26 with an assumption of no signal interference with any nearby wireless devices, motes are located within the 2.4 GHZ ISM band communication range with full transmission power of 0dBm, the sensing mote is equipped with new full power 3.0V AA Alkaline batteries, and both motes are not contained in any container for weather and environment protection purposes.

C. Software Settings

In this experiment, we programmed the TelosB motes using the TinyOS 2.1.1 nesC codes running on a 32-bit Ubuntu 14 OS. The Ubuntu Linux is running on the Parallels Desktop virtual machine version 11.1.3 on an OS X EI Capitan version 10.11.3 computer.

D. Systems Wiring

In nesC, the connection among components involved in the systems is called '*wiring*'. nesC codes wiring occurs when connecting between the interface of the *user* that used the component to the *provider* of the interface. In other words, the interface's *user* is connected to or make requests (by *calling commands*) on the interface's *provider* (the *provider* makes callbacks to the interface's *user* by *signaling events*) [16], [17], [18].

In our case of reading the temperature, the sensing application component (called *module*) is a user of the *Read* interface (Read<uint16_t>) provided by the *SensirionSht11C* component (called *configuration*); where the *readDone* event is a callback signaled when the sensor component finished sensing the temperature. The wiring between these components are shown below:

```
components new SensirionSht11C() as TempSensor;
TempSenseRadioC.Read -> TempSensor.Temperature;
```

The operator '->' infers that the TempSenseRadioC module uses the *Read* interface provided by the SensirionSht11C components (as the TempSensor instance). Here, TempSenseRadioC is the *user* while SensirionSht11C is the *provider*. The '->' operator also denotes the *wiring* between the TempSenseRadioC module and the SensirionSht11C components. The definition of our sensing application module in the nesC program code is:

```
module TempSenseRadioC {
    uses {
        interface Read<uint16_t>;
    }
}
```

A call to the interface Read is done by calling the *read()* function provided by the SensirionSht11C component. The function call denotes by: *call Read.read()*. We set this function call in an event of a timer being fired (*event void MilliTimer.fired()*), which is at every one second; each second the timer fired, the SensirionSht11C sensor component on the TelosB mote sense the temperature in the area of sensing.

As a user of the *Read* interface, the TempSenseRadioC module also must implement any events issued by the provided components (SensirionSht11C), where in this case is the readDone event and denotes by the below code.

```
event void Read.readDone(error_t result, uint16_t data)
```

While receiving this signal callbacks, we program our sensing application to read the message structure of the received temperature readings (of a size of 14-bit readings read in an unsigned 16-bit data type), calibrate the readings based on the calibration formula stated in the sensor datasheet, and display the meaningful readings on the base station computer screen (shown in next section).

In particular, our sensing application module is named TempSenseRadioC while the configuration application that wire the TempSenseRadioC to the TinyOS services is named TempSenseRadioAppC. The TempSenseRadioAppC configuration are details as below:

```
configuration TempSenseRadioAppC {}
 implementation {
   components MainC, LedsC, TempSenseRadioC;
   components new AMSenderC(AM_SENSE_MSG);
   components new AMReceiverC(AM_SENSE_MSG);
   components new TimerMilliC();
   components ActiveMessageC;
   components new SensirionSht11C() as TempSensor;

   TempSenseRadioC.Boot -> MainC.Boot;
   TempSenseRadioC.Read -> TempSensor.Read;
   …
   …
}
```

This denotes that the TempSenseRadioAppC configuration is built out of several components (modules or configurations) namely, MainC (system boot), LedsC (LED control), TempSenseRadioC (our sensing module), AMSenderC (radio control), AMReceiverC (radio control), TimerMilliC (timer control), ActiveMessageC (TinyOS's messaging system), and SensirionSht11C (temperature sensor control). TempSenseRadioAppC explicitly specifies the connection (or *wiring*) between interface provided and used by these components. Fig. 1 demonstrate the wiring diagram of our TempSenseRadioC sensing application.

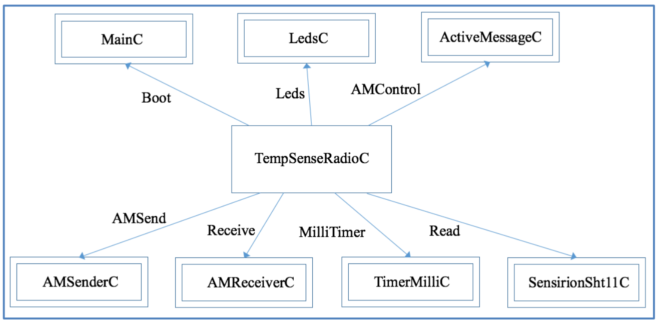


Fig. 1 Wiring diagram for TempSenseRadio application. The connecting arrows show Interfaces from the *use* module (TempSenseRadio) to the *provider* components.

E. System Configuration

The objective of the experiment is to send temperature readings using wireless transmission from the sensing mote to the base station, where the connection between the sensing mote to the base station is via 2.4 GHz radio and the connection between the base station and the attached computer is using a USB connector. Fig. 2 shows the connection setup.

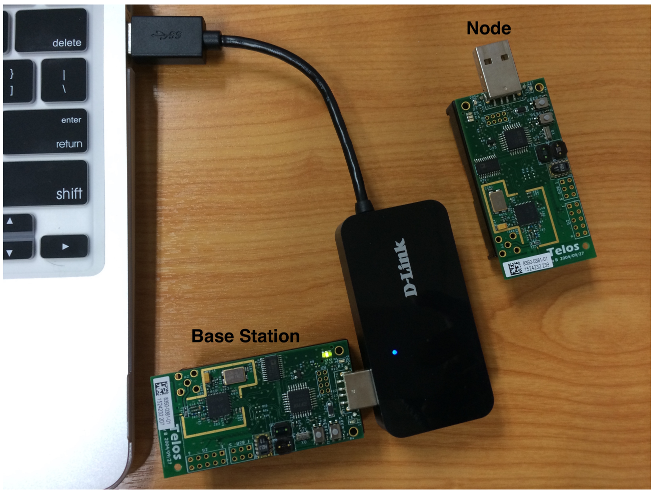


Fig. 2 Logical connection of the sensing mote (node) and the base station mote. The sensing mote connects to the base station using 2.4 GHz radiowave while the base station connects to the monitoring computer using USB-to-serial port.

We built the sensing application in TinyOS using nesC codes. The application is installed on the sensing mote to sense the temperature of a room. Each time the mote done with each of the sensing, it triggers its led component (the yellow led) to blink, as an indicator to the observer that data has being sensed at that particular period of time. We set the periodic sensing task to be at every 1024 millitime or equivalent to 1 second.

As soon as the mote boots up, it powers on the radio component to begin the transmission. This is also the time the periodic timer starts ticking its clock for periodic data sensing. As been indicated, the sensing mote Timer keeps ticking for every second which also indicates the mote to be sensing the temperature readings for every second. The nesC codes to start the radio as the mote booted up and starting the periodic timer are as follows:

```
event void Boot.booted() {
    call AMControl.start();
}

event void AMControl.startDone(error_t err) {
    if (err == SUCCESS) {
        call MilliTimer.startPeriodic(1024);
    }else {
        call AMControl.start();
    }
}
```

Similarly, the yellow led (*led1*) of the mote is blinking (toggling) synchronically with the sensing of the temperature, and at the same time the mote is sending the sensed data to the base station through its radio transceiver.

```
event void MilliTimer.fired() {
    sense_count_msg_t* scm = (sense_count_msg_t*)call
    Packet.getPayload(&packet, sizeof(sense_count_msg_t));
    call Read.read();
    call Leds.led1Toggle();
    if (scm == NULL) {
        return;
    }
    if (call AMSend.send(AM_BROADCAST_ADDR,
    &packet, sizeof(sense_count_msg_t)) == SUCCESS) {}
}
```

In short, as the radio started, three events occur: periodic timer starts ticking, data being sensed, and the sensed data is transmitted to the base station. All these events occur once at a time in one clock tick.

The sensing task consists of two processes. First, the data sensing that occurs at every one second during the clock tick. Second task occurs particularly after the sensing finished. During this time, the sensed data are read and stored in the packet structure before transmitting to the base station. It is actually during at the second task, the raw sensed data are interpreted, processed, and the readings are calibrated before it can understandably be viewed to the human eyes.

Calibrating the sensor reading during the second process of the read event, need to apply the formula given from the developer specification manual. For example, to convert a 14-bit temperature reading with mote power of 3.0 Volt to temperature Celsius and Fahrenheit, we use the following formula [15]:

$$T_C = -39.6 + 0.01 * T_R$$

$$T_F = -39.3 + 0.018 * T_R$$

where T_C is temperature in Celcius, T_F is temperature in Fahrenheit, and T_R is the sensed temperature. In order to view the resulted temperature on the screen, we make use of the *printf* function available from the *printf.h* library provided by TinyOS. Since the temperature data structure is of a type of unsigned integer (uint16_t), we have to use the '%u' operator to correctly print and view the calibrated temperature. The snippet codes below show the nesC codes of these events.

```
event void Read.readDone(error_t result, uint16_t data) {
    uint16_t tempCelc;
    uint16_t tempFah;

    sense_count_msg_t* scm = (sense_count_msg_t*)call
    Packet.getPayload(&packet, sizeof(sense_count_msg_t));

    if (result == !SUCCESS) {
        data = 0xffff;
    }else{
        scm->tempReading = data;
        tempCelc = -39.6+0.01*scm->tempReading;
        tempFah = -39.3+0.018*scm->tempReading;

        printf("Reading: 0x%x (Hex) / %u (Dec)\n",
        scm->tempReading, scm->tempReading);
        printf("Temperature: %u (C) / %u (F)\n",tempCelc,
        tempFah);
        printfflush();
    }
}
```

To view the temperature on the computer screen, we simply use the built-in java tool of TinyOS called *PrintfClient* by hitting a command below. The sample output of the readings is shown in Fig. 3.

$java net.tinyos.tools.PrintfClient –comm serial@/dev/ttyUSB*x*:telosb

```
Thread[Thread-1,5,main]serial@/dev/ttyUSB1:115200: resynchronising
Reading: 0x19a5 (Hex) / 6565 (Dec)
Temperature: 26 (C) / 78 (F)
Reading: 0x19a2 (Hex) / 6562 (Dec)
Temperature: 26 (C) / 78 (F)
Reading: 0x19a2 (Hex) / 6562 (Dec)
Temperature: 26 (C) / 78 (F)
Reading: 0x19a1 (Hex) / 6561 (Dec)
Temperature: 26 (C) / 78 (F)
Reading: 0x199e (Hex) / 6558 (Dec)
Temperature: 25 (C) / 78 (F)
Reading: 0x199d (Hex) / 6557 (Dec)
Temperature: 25 (C) / 78 (F)
Reading: 0x199c (Hex) / 6556 (Dec)
Temperature: 25 (C) / 78 (F)
Reading: 0x199b (Hex) / 6555 (Dec)
Temperature: 25 (C) / 78 (F)
```

Fig. 3 Temperature sensor readings in Celcius (C) and the equivalent Fahrenheit (F) after applying the calibration formula.

Part of the experiment also dealt with the hopping of the sensed readings from the sensing mote to the base station. Act as a ‘communicating bridge’, the base station relays any serial or radio data it receives via its serial port or radio transceiver to the attached computer. In our experiment, we view the readings in a form of bytes containing the packet header together with the sensor readings. We use the TinyOS’s built-in ‘*Listen*’ java tool by issuing $java net.tinyos.tools.Listen –comm serial@/dev/ttyUSB*x*:telosb to display these data. Fig. 4 shows the resulting screen output.

```
serial@/dev/ttyUSB0:115200: resynchronising
00 FF FF 00 01 02 00 06 19 6D
00 FF FF 00 01 02 00 06 19 6D
00 FF FF 00 01 02 00 06 19 6E
00 FF FF 00 01 02 00 06 19 6E
00 FF FF 00 01 02 00 06 19 6E
00 FF FF 00 01 02 00 06 19 6D
00 FF FF 00 01 02 00 06 19 6E
00 FF FF 00 01 02 00 06 19 6D
00 FF FF 00 01 02 00 06 19 6C
00 FF FF 00 01 02 00 06 19 6D
00 FF FF 00 01 02 00 06 19 6D
00 FF FF 00 01 02 00 06 19 6E
00 FF FF 00 01 02 00 06 19 6C
```

Fig. 4 Temperature sensor readings (two bytes from right) in raw byte format shown by issuing the TinyOS’s *Listen* java tool at the base station mote.

The output shows a complete TinyOS’s ActiveMessage packet with the payload of the packet is at the last two bytes of the packet (*little endian* notation), the rest are the packet header. Therefore, by examining each line (packet) of the output, the temperature readings are byte number shown as ’19 6D’, ’19 6E’, ’19 6C’, etcetera. By translating to the Celcius / Fahrenheit equivalent, these readings represent a temperature of 25°C / 77°F.

In conclusion, the sensor readings can be viewed either using the *PrintfClient* or the *Listen* java tool, with a simple calibration on the received readings. To aid the translation of the readings using the *Listen* tool, we had created a simple C programs to display similar output as shown in Fig. 3. Fig. 5 shows a snapshot of the main function of the program.

```c
void main(){

    unsigned long tempDec;
    signed long tempCelc;
    signed long tempFah;
    char hex[4];//4 characters for 16-bit Hexadecimal

    printf("Enter temperature reading (16-bit hex): ");
    scanf("%s",hex);

    tempDec = convHexToDec(hex);
    /*
    Sensirion SHT11 Temp calibration formula: T = d1 + d2 * SOt
    VDD:3V, SOt=14bit -> d1=-39.6(C),-39.3(F)  d2=0.01(C),0.018(F)
    */
    tempCelc = -39.6+0.01*tempDec; //Celcius
    tempFah = -39.3+0.018*tempDec; //Fahrenheit

    printf("Reading: %s (Hex), %lu (Dec)\n", (char*)hex, tempDec);
    printf("Temperature: %ld (C) / %ld (F)\n",tempCelc, tempFah);
}
```

Fig. 5 The C main program to calibrate and convert the TelosB’s Sensirion SHT11 sensor readings to their Celsius and Fahrenheit equivalent.

IV. Conclusion and Future Works

In this study, we demonstrate the application of environmental sensing and the basic science of data sensing, data processing, and data display using the TinyOS TelosB motes. The experiment setup is simple: sensing temperature using one mote, calibrate the sensed data, hop the data to the base station, and display them on the screen for monitoring. The significance of the study is twofold: the basic knowledge on sensor data capturing we gather in this experiment is valuable to open up a further research on integrating the IPv6, 6LoWPAN, and CoAP protocol in order to view the sensed data globally through different networks in the World Wide Web; and the basic knowledge we gather on hopping the sensed data through the radio communication would eventually important for us to research on the energy-efficient techniques of data passing from one node to another in a networked multihopping fashion.

It is therefore, our future works may reserve to investigate the integration of the Internet of Things technology in sensor data monitoring, and the energy-efficiency mechanisms on information passing in the networked of sensors.

Acknowledgment

This work is partially supported by the Ministry of Education Malaysia and the Encouragement Grant with Reference No. Q.K130000.2638.11J47 of Universiti Teknologi Malaysia (UTM).

References

[1] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, “A survey on sensor networks,” *IEEE Commun. Mag.*, vol. 40, no. 8, pp. 102–105, 2002.

[2] S. R. Wankhade and N. A. Chavhan, “A review on data collection method with sink node in wireless sensor networks,” *Int. J. Distrib. Parallel Syst.*, vol. 4, no. 1, pp. 67–74, 2013.

[3] R. Gao, H. Zhou, and G. Su, “A wireless sensor network environment monitoring system based on TinyOS,” *2011 Int. Conf. Electron. Optoelectron. (ICEOE 2011)*, pp. 497–501, 2011.

[4] B. Krishnamachari, *Networking wireless sensors*. Cambridge: Cambridge University Press, 2005.

[5] J. Barros, "Sensor networks: An overview," in *Learning from data streams*, New York, USA: Springer Berlin Heidelberg, 2007, pp. 9–22.

[6] H. Karl and A. Willig, *Protocols and architectures for wireless sensor networks*. West Sussex, England: John Wiley & Sons, 2007.

[7] P. Sanyal, S. Das, S. S. Bhunia, S. Roy, and N. Mukherjee, "An experience of implementing IPv6 based data retrieval system for Wireless Sensor Networks," *2012 Int. Conf. Recent Adv. Comput. Softw. Syst.*, pp. 154–157, 2012.

[8] F. Wang and J. Liu, "Networked wireless sensor data collection: issues, challenges, and approaches," *IEEE Commun. Surv. Tutorials*, vol. 13, no. 4, pp. 673–687, 2011.

[9] IEEE Standard, "IEEE standard for information technology-telecommunications and information exchange between systems-local and metropolitan area networks-specific requirements-Part 15 . 4 : Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications ," *IEEE Std 802.15.4-2006*, pp. 1–26, 2006.

[10] F. Zhao and L. Guibas, *Wireless sensor networks: An information processing approach*. San Jose, CA, California: Elsevier, 2004.

[11] TinyOS, "Mote-mote radio communication." [Online]. Available: http://tinyos.stanford.edu/tinyos-wiki/index.php/Mote-mote_radio_communication. [Accessed: 20-Jun-2003].

[12] MEMSIC Inc., "TelosB datasheet: Document Part Number: 6020-0094-02 Rev B." San Jose, CA, California, 2003.

[13] T. Instruments, "CC2420 datasheet: 2.4 GHz IEEE 802.15.4/ZigBee-ready RF transceiver: Document SWRS041B." Dallas, Texas, pp. 1–89, 2010.

[14] J. Polastre, R. Szewczyk, and D. Culler, "Telos: Enabling ultra-low power wireless research," *2005 4th Int. Symp. Inf. Process. Sens. Networks, IPSN 2005*, pp. 364–369, 2005.

[15] Sensirion, "Datasheet SHT1x (SHT10, SHT11, SHT15) Humidity and Temperature Sensor," no. September. pp. 1–11, 2008.

[16] D. Gay, P. Levis, R. von Behren, M. Welsh, E. Brewer, and D. Culler, "The nesC language : A holistic approach to networked embedded systems," *Proc. ACM SIGPLAN 2003 Conf. Program. Lang. Des. Implement.*, vol. 38, no. 5, pp. 1–11, 2003.

[17] P. Levis and D. Gay, *TinyOS Programming*, vol. 1st. Ed. 2009.

[18] P. Levis, S. Madden, J. Polastre, R. Szewczyk, K. Whitehouse, A. Woo, D. Gay, J. Hill, M. Welsh, E. Brewer, and D. Culler, "TinyOS: An operating system for wireless sensor networks," in *Ambient Intelligence*, Springer-Verlag, 2005, pp. 115–148.

Compressed Image Transmission over AWGN Channel using DCT and Raised Cosine Filter

Md. Khaliluzzaman*
Dept. of Computer Science & Engineering (CSE)
International Islamic University Chittagong (IIUC)
Chittagong-4203, Bangladesh
*khalil_021@yahoo.co.in

Deepak Kumar Chy.
Dept. of Electrical & Electronics Engineering (EEE)
University of Information Technology & Sciences (UITS)
Dhaka-1212, Bangladesh
dk_chy53@yahoo.de

Abstract - Compressed image transmission through model channel like AWGN implementing DCT and Raised Cosine Filter is demonstrated in this paper. The transmission performance of still image with Qam-32 modulation scheme is compared to communication model having no filter in presence of additive white Gaussian noise (AWGN). By comparing communication model having filter in the same scenario, results showed that the received image quality is better in terms of PSNR and MSE for low value of SNR. The simulation results reveal that with lower signal to noise ratio, raised cosine filter exhibits better performance in comparison with employing no filter. Inter symbol Interference (ISI) that plays an important role in image transmission degrades the image quality and may be eliminated by employing filter.

Keywords- AWGN; DCT; Raised Cosine Filter; QAM; PSNR; MSE; ISI; SNR

I. INTRODUCTION

During the past few years, people begin to play more and more attention on reliable wireless multimedia transmission, such as Image, Video, and Audio based on high speed data communication, high quality channel and high quality transmission of visual data, for instance, high quality image over the communication channel. In general, High quality image transmission demands more storage capacity and high bandwidth requirements. For that reason, image compression is performed so that less storage capacity and required bandwidth is achieved. Nevertheless, compressed image is prone to noise both in wire and wireless communication channel. In addition, ISI degrade the image quality over wireless channel and /or wired channel due to band limited characteristic of transmitted data. Appropriate filter technique may improve the image quality that transmitted through the noisy channel.

In this paper, a raised cosine filter is proposed before transmission of compressed image through noisy communication channel. Typical communication model is tested by transmitting still image using conventional QAM -32 modulation scheme. The simulation results proved that proposed communication model for image transmission yields quality image when channel is more vulnerable to noise.

In this section provide a descriptive summary of some technique that have been implemented and tested for image compression and transmission over noisy channels. A. Mishra et al. [1] proposed the polar coding for gray scale image transmission over AWGN channel using QAM-64. However, does not suggest about band limited characteristic of medium and the influence of inter symbol interference. On the other hand, Md. Abdul Kader et al. [2] suggested the image transmission using Hierarchical Quadrature Amplitude Modulation (HQAM) for better protection of high priority data. Nevertheless, no channel model is considered such as AWGN and salt and pepper noise has been taken into account. Furthermore, Md. A. H. Khandokar et al. [3] used a simple conventional communication model with various M – array modulation technique with AWGN Channel but does not show the lower value of E_b/N_o. Only the value of 10 dB has been used. This paper has investigated the conventional communication model with proposed communication model for image transmission using filter and the effect of low and high signal to noise ratio have been considered for both communication models for still image transmission.

The paper is organized as follows. In Section II, theoretical background of DCT compression, 32-QAM modulation, raised cosign filter, White Gaussian Noise (AWGN) channel, signal to noise ratio (SNR), bit error rate (BER), and E_b/N_0 (Energy per bit to Noise power spectral density ratio) has been demonstrated. In Section III the proposed method is described. In the next section the experimental result is explained. The paper is concluded in Section IV.

II. THEORETICAL BACKGROUND

A. DCT Compression

The proposed compression method is based on the Discrete Cosine Transform (DCT) [15] applied to the global image. The complete image is considered as a single block. Let I be the original image defined as I=If (u,v), {O<u$\leq M-1$, 0<v$\leq N-1$}. MxN is the dimension of the image and A (u,v) denote the gray level pixel's at (u, v) coordinates. The MxN DCT coefficients are given by :

$$F(i,j) = \frac{2c(i)c(j)}{\sqrt{MN}} \sum_{u=0}^{M-1} \sum_{v=0}^{N-1} f(u,v) \cos\left[\frac{(2u+1)i\pi}{2M}\right] \cos\left[\frac{(2v+1)j\pi}{2N}\right] \quad (1)$$

where, c(:) is define by.

$$c(i) = \begin{cases} \frac{1}{\sqrt{2}}, & if\ i = 0 \\ 1, & if\ i \neq 0 \end{cases}$$

Respectively, the inverse DCT coefficients are given by:

$$f(u,v) = \sum_{u=0}^{M-1} \sum_{v=0}^{N-1} \frac{2c(i)c(j)}{\sqrt{MN}} f(i,j) \cos\left[\frac{(2u+1)i\pi}{2M}\right] \cos\left[\frac{(2v+1)j\pi}{2N}\right] \quad (2)$$

In the following sections, we consider square images (M=N).

B. 32-QAM modulation technique

QAM is used in applications including microwave digital radio, DVB-C (Digital Video Broadcasting Cable), and so on. In 32-QAM, Quadrature Amplitude Modulation used in this paper has five I values and five Q values. This results in a total of 32 possible states for the signal. It can transition from any state to any other state at every symbol time. Since $32 = 2^5$, five bits per symbol can be sent which is as shown in Fig. 1. This is too many states for a power of two (the closest power of two is 32). So the four corner symbol states, which take the most power to transmit, are omitted. This consists of two bits for I and two bits for Q. The symbol rate is one fifth of the bit rate. So this modulation format produces a more spectrally efficient transmission. It is more efficient than BPSK, QPSK, or 8PSK. However, the symbols are very close together and are thus more prone to errors due to noise and distortion which is shown in simulation result [5].

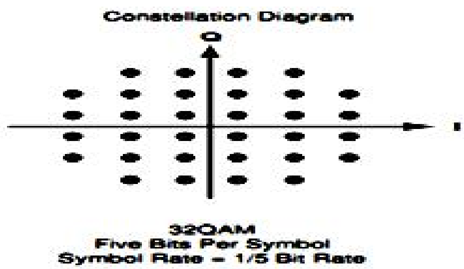


Fig. 1. Constellation diagram for QAM-32.

C. FIR raised cosine filter

In modern data transmission systems, Image is transmitted as bits or groups of bits (symbols) typically in the form of individual pulses of energy. Sometimes rectangular pulse is probably the most fundamental. It is easy to implement in a real-world system because it can be directly compared to opening and closing a switch, which is synonymous with the concept of binary information. Pulses sent by the transmitter are detected by the receiver in any data transmission system. At the receiver, the goal is to sample the received signal at an optimal point in the pulse interval by the matched filter to maximize the probability of correct decision. This implies that the fundamental shapes of the pulses be such that they do not interfere with one another at the optimal sampling point. There are two criteria that ensure non-interference [7].

a. The pulse shape exhibits a zero crossing at the sampling point of all pulse intervals except its own. That is Minimized inter symbol interferences (ISI).

b. The shape of the pulses is such that the amplitude decays rapidly outside of the pulse interval. That is high stop band attenuation.

The rectangular pulse meets first requirement because it is zero at all points outside of the present pulse interval. It cannot cause interference during the sampling time of other pulses. The trouble with the rectangular pulse, however, is that it has significant energy over a fairly large bandwidth as indicated. The unbounded frequency response of the rectangular pulse makes it unsuitable for modern transmission systems. This is where pulse shaping filters come into play. If the rectangular pulse is not the best choice for band-limited data transmission, then what pulse shape will, decay quickly, and provide zero crossings at the pulse sampling times [5]. There are several choices that have but in most systems Raised cosine filter are used to shape the input pulse.

In most cases, the square root raised cosine filter is used in the transmitter and receiver part of the system so that the overall response resembles that of a raised cosine filter. The impulse or time domain response of the raised cosine filter and the square root raised cosine filter are given by (3), (4), (5), (6) [7].

$$h_{RC}(t) = \frac{\sin(\frac{\pi T}{T})}{(\frac{\pi T}{T})} \frac{\cos(\frac{\pi \alpha T}{T})}{1-(\frac{\pi T}{T})^2} \tag{3}$$

This expression can be simplified further by introducing the sinc function ($sincx = \frac{\text{Sin x}}{\text{x}}$).

$$h_{RC}(t) = \sin c\,(\frac{\pi T}{T}) \frac{\cos(\frac{\pi \alpha T}{T})}{1-(\frac{\pi T}{T})^2} \tag{4}$$

The sinc function in the response of the filter ensures that the signal is band-limited. The time domain or impulse response of the square root raised cosine filter is given as;

$$h_{RRC}(t) = \frac{\sin[\pi(1-\alpha)t]+4\alpha\left(\frac{t}{T}\right)\cos[\pi(1+\alpha)\frac{t}{T}]}{\left(\frac{\pi T}{T}\right)[1-(\frac{4\alpha t}{T})^2]} \tag{5}$$

The overall response of the system is given by (14).

$$h_{RC}(t) = h_{RRC}(t)\, h_{RRC}(t) \tag{6}$$

The impulse and magnitude response of Raised Cosine filter are shown in Fig. 2 and Fig. 3.

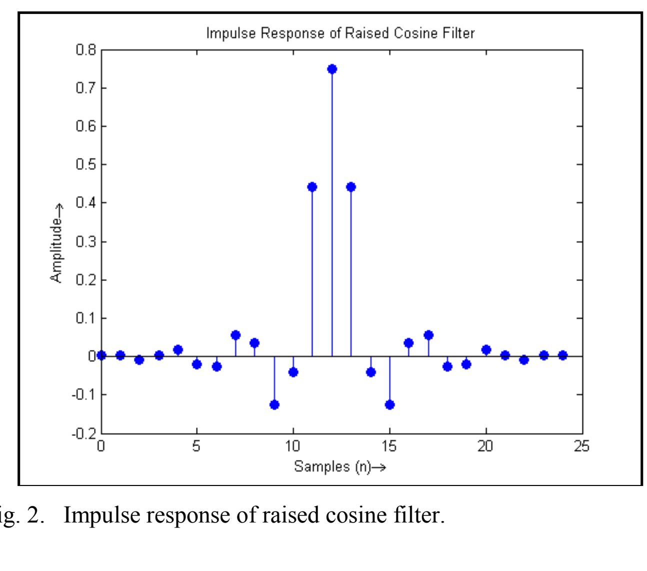


Fig. 2. Impulse response of raised cosine filter.

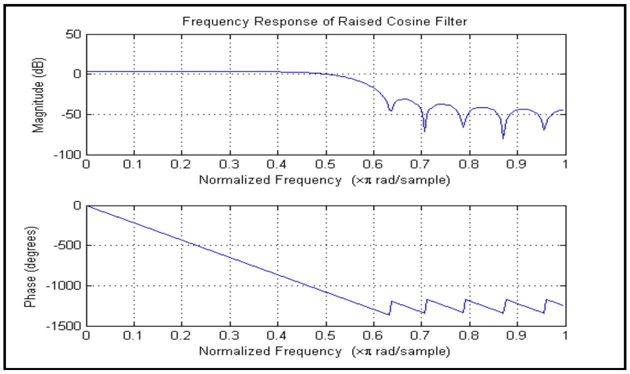


Fig. 3. Frequency response of raised cosine filter.

D. *White Gaussian Noise (AWGN) channel*

It is modeled as a zero-mean Gaussian random process where the random signal is the summation of the random noise variable and a direct current signal as shown in (7) [11], [12] and [13].

$$Z = a + n \tag{7}$$

The probability distribution function for this Gaussian noise can be represented as:

$$p(z) = \frac{1}{\sigma\sqrt{2\pi}}\exp[-\frac{1}{2}(\frac{z-a}{\sigma})^2] \tag{8}$$

The model of this noise assumes a power spectral density $G_n(f)$ which is flat for all the frequencies denoted as;

$$Gn\,(f) = \frac{N_0}{2} \tag{9}$$

The factor 2 indicates that the power spectral density is a two-sided spectrum. This type of noise is present in all communication systems and is the major noise source for most systems with characteristics of additive, white and Gaussian. It is mostly used to model noise in communication systems which are simulated to determine their performance. This noise is normally used to model digital communication systems which can be replaced with other interference schemes.

E. *signal to noise ratio (SNR)*

SNR is the ratio of the received signal strength over the noise strength in the frequency range of the operation. It is an important parameter of the physical layer of Local Area Wireless Network (LAWN). Noise strength, in general, can include the noise in the environment and other unwanted signals (interference). BER is inversely related to SNR, that is high BER causes low SNR. High BER causes increases packet loss, increase in delay and decreases throughput [5]. The exact relation between the SNR and the BER is not easy to determine in the multi channel environment. Signal to noise ratio (SNR) is an indicator commonly used to evaluate the quality of a communication link and measured in decibels and represented by (10).

$$\text{SNR} = 10\log10\,(\text{Signal Power} / \text{Noise Power})\ \text{dB} \tag{10}$$

F. *Bit error rate (BER)*

The BER, or quality of the digital link, is calculated from the number of bits received in error divided by the number of bits transmitted.

$$\text{BER} = (\text{Bits in Error}) / (\text{Total bits received}) \tag{11}$$

In digital transmission, the number of bit errors is the number of received bits of a data stream over a communication channel that has been altered due to noise, interference, distortion or bit synchronization errors. The BER is the number of bit errors divided by the total number of transferred bits during a particular time interval. BER is a unit less performance measure, often expressed as a percentage [10].

BER can also be defined in terms of the probability of error (POE) [14] and represented by (12).

$$\text{POE} = \frac{1}{2}\,(1 - \text{erf})\sqrt{\frac{E_b}{N_0}} \tag{12}$$

Here, ‘erf’ is the error function, E_b is the energy in one bit and N_0 is the noise power spectral density (noise power in a 1Hz bandwidth). The error function is different for the each of the various modulation methods. The POE is a proportional to E_b/N_0, which is a form of signal-to-noise ratio. The energy per bit, E_b, can be determined by dividing the carrier power by the bit rate. As an energy measure, E_b has the unit of joules. N_0 is in power that is joules per second, so, E_b/N_0 is a dimensionless term, or is a numerical ratio.

G. *E_b/N_0 (Energy per bit to Noise power spectral density ratio)*

E_b/N_0 is an important parameter in digital communication or data transmission. It is a normalized signal-to- noise ratio (SNR) measure, also known as the "SNR per bit". It is especially useful when comparing the bit error rate (BER) performance of different digital modulation schemes without Taking bandwidth into account. *E_b/N_0* is equal to the SNR divided by the "gross" link spectral efficiency in (bit/s)/Hz, where the bits in this context are transmitted data bits, inclusive of error correction information and other protocol overhead.

III. Proposed Method

This section has described the proposed method in details. The basic steps of the proposed method are, 1) converting RGB image into gray scale image, 2) applying DCT in gray scale image to compress and quantized the gray scale image, 3) then encoded the quantized image and finally find the compressed image, 4) utilizing modulation technique on encode values to obtain the symbols, 5) before transmitting symbol value through AWGN channel, the symbol data values are filtered with raised cosine filter and apply the Inverse first fourier transform to convert the value from frequency domain to time domain, 6) at receiver site the noisy values are passed through the inverse raised cosine filter and then converted to time domain to frequency domain by applying fourier transform, 7) these transformed values are demodulated with QAM 32 and, 8) decode the demodulated values accordingly, 9) after that, the

IDCT has been applied to get the original gray scale values, and 10) finally, retrieve the original image. The workflow of the proposed framework is as shown in Fig. 4.

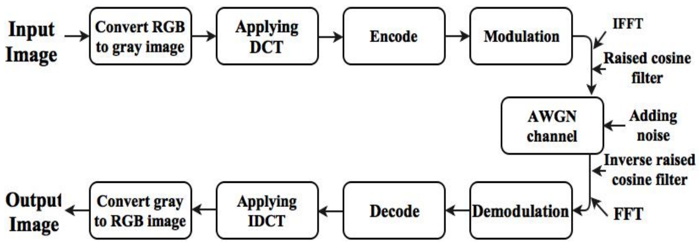


Fig. 4. Workflow of the proposed framework.

A. Converting RGB image into gray scale image

In this section, initially the input RGB image is converted into gray scale image. The gray scale image contains only one channel value that is intensity value. The RGB image is converted to the gray scale image by using the formula mentioned in (13).

$$I_{gray} = 0.299*R + 0.587*G + 0.114*B \tag{13}$$

B. Applying DCT to compress the gray scale image

This section utilizes the image that is converted to the gray scale in the previous section. In an image, low frequency values are accumulated at the left upper corner and high frequency values are stored at lower right corner of the compressed image block. As most of the valuable information is stored at low frequencies, discarding certain information from higher frequency has little effect on the overall image quality.

Typically 8x8 or 16x16 block size is used instead of implementing DCT on the entire image. The DCT transform is usually applied firstly in row wise direction, then column wise direction. Avoiding the complexity, column wise DCT operation can be accomplished applying row wise DCT operator. For this reason, at first row wise DCT operation is applied on the image block, then transpose is implemented on the column wise and then perform the row wise DCT operation once again. The procedure is as shown in Fig. 5.

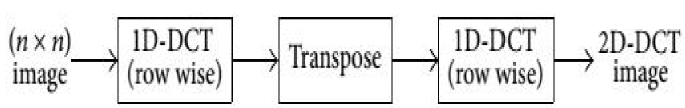


Fig. 5. Two dimensional DCT operation.

DCT operation has high degree of computational complexity. The mathematical expression for performing N-point DCT operation is given in (1). For instance, 8-point DCT operation equals 8 in (1). The inputs to the *8-point DCT* operator are eight pixel values (u(0).....u(7)). After executing the DCT computation, eight DCT values (F(0)......F(7)) are yield at the operator output. For calculating each DCT value, the all obtained input values such as (u(0).....u(7)) are used. These values are finally quantized. Quantization is achieved by dividing transformed image matrix by the quantization matrix used. Values of the resultant matrix are then rounded off.

C. Encode

In this section, the compressed quantization values are encoded to convert into binary code streams. The most commonly used entropy encoders are the Huffman encoder and the arithmetic encoder, although for applications requiring fast execution, simple run-length encoding (RLE) has proven very effective.

D. Utilizing 32-QAM modulation

In this section, 32-QAM, Quadrature Amplitude Modulation is used in this paper which has five I values and five Q values. This results in a total of 32 possible states for the signal. It can transit from any state to any other state at every symbol time. Since $32 = 2^5$, five bits per symbol can be sent. This is too many states for a power of two (the closest power of two is 32). So the four corner symbol states, which take the most power to transmit, are omitted.

Here 32-QAM is used as there is a tradeoff between power efficiency and bit error rate.

E. Applying raised cosine filter and IFFT

This section uses the root raised cosine filter to eliminate inter symbol interference which degrade the image quality severely at low signal to nose ratio. Single raised cosine filter at transmitter side cannot reduce the inter symbol interference. Therefore, a pair of raised cosine filter is used in this paper which acts as root raised cosine filter. The impulse and frequency response of the raised cosine filter are shown in Fig. 2 and Fig. 3. In addition, as original image block is in frequency domain but raised cosine filter and AWGN model work on real time operation. For that purpose, inverse first Fourier transform (IFFT) is introduced in this proposed method before using raised cosine filter.

F. AWGN channel

In practice, according to Shannon's capacity theorem, there is no channel which is noise free. All transmitted signals are corrupted by noise and noise is unpredictable in nature. Precise mathematical operations can't be done on noise. That's why AWGN channel is chosen as channel model for simulation.

G. Retrieving original image

In this section, how original image is retrieved at receiver side is demonstrated. Firstly, the noisy image data are passed through the inverse raised cosine filter to obtain the modulated image data in time domain. Then, demodulation process is applied followed by first Fourier transform to obtain the encoded data in frequency domain. After that, the decoding technique is implemented to obtain quantized image value. In addition to this, inverse discrete cosine transform (IDCT) operation is performed to obtain the gray scale vale. Finally, the original RGB image is retrieve from gray scale image by using appropriate method.

IV. SIMULATION AND EXPERIMENTAL RESULTS

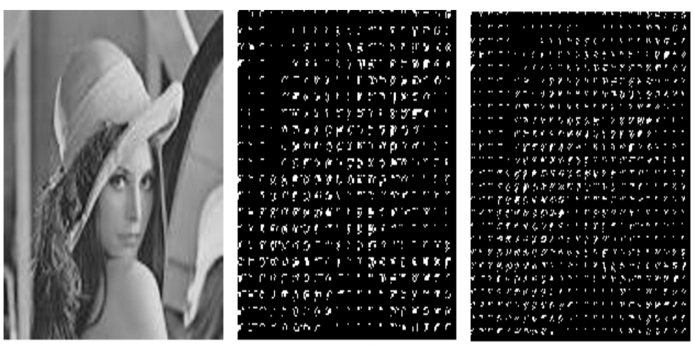
In this paper, simulation is carried out for gray scale images. The simulation was performed using MATLAB environment.

The mean square error (MSC) and peak signal to noise ratio (PSNR) are typically used to measure the quality of the receiving image with respect to transmitting image. The MSE and PSNR are measured using (14) and (15).

$$MSE = \frac{1}{MN}\sum_{i=0}^{M-1}\sum_{j=0}^{N-1}\|I(i,j) - R(i,j)\|^2 \quad (14)$$

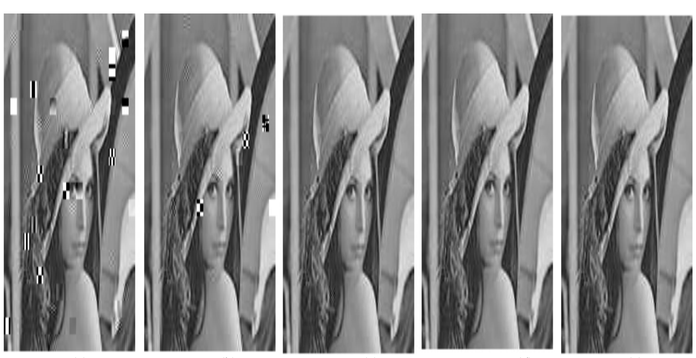
$$PSNR = 10log_{10}^{\frac{\|I(i,j)-R(i,j)\|^2}{MSE}} \quad (15)$$

The result of compressed quantized image for the input image with and without filter is shown in Fig. 6 (b) and Fig. 6(c) respectively. Fig. 7 shows the processing example of retrieve images with E_b/N_0(dB) effect where raised cosine filter is used. From the experimental results, it is seen that the received image quality has improved with the increase of E_b/N_0(dB) values. When the value of E_b/N_0(dB) is greater than five the bit error rate is zero and the value of MSE and PSNR is constant which is 23.5973 and 4.1420 respectively. The result of bit error rate, MSE, and PSNR with respect to E_b/N_0(dB) is as shown in Table I, where raised cosine filter was present.



(a) (b) (c)

Fig. 6. Processing example of quantized image: a) original image, b) quantized image using filter, and c) quantized image without filter.



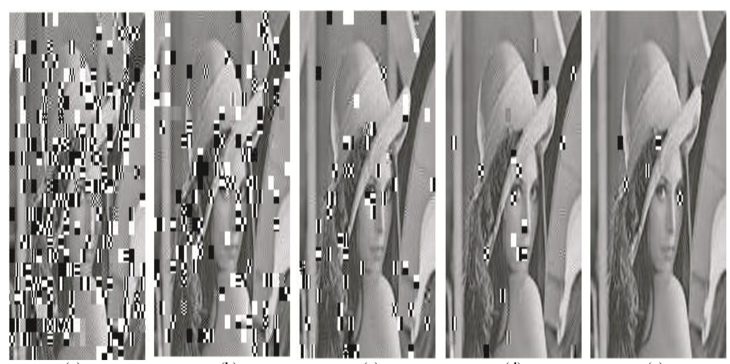
(a) (b) (c) (d) (e)

Fig. 7. Processing example of output images with different E_b/N_0(dB) value using raised cosine filter: a) E_b/N_0(dB)=1, b) E_b/N_0(dB)=2, c) E_b/N_0(dB)=3, d) E_b/N_0(dB)=4, and e) E_b/N_0(dB)=5

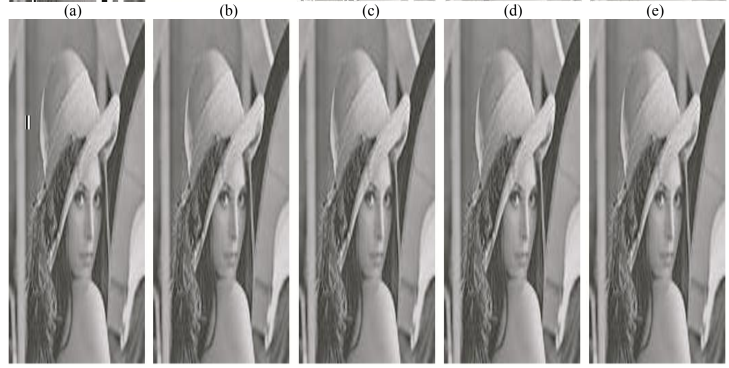
TABLE I. EXPERIMENTAL RESULT OF NO OF ERROR, BER, MSC AND PSNR USING WITH FILTER

E_b/N_0(dB)	Number of error	Bit error rate(BER)	MSC	PSNR
1	1625	$9.9182xe^{-4}$	NA	NA
2	467	$2.8503xe^{-4}$	NA	NA
3	68	$4.1504xe^{-5}$	25.1261	3.8900
4	9	$5.4932xe^{-6}$	23.5973	4.1420
5	1	$6.1035xe^{-7}$	23.5973	4.1420
6	0	0	23.5973	4.1420
7	0	0	23.5973	4.1420
8	0	0	23.5973	4.1420
9	0	0	23.5973	4.1420
10	0	0	23.5973	4.1420

Fig. 8 shows the processing example of retrieving images with E_b/N_0(dB) effect where raised cosine filter was not used. The experimental result shows that with the lower E_b/N_0(dB) value the received image without filter is more blurred with respect to the images that are received with filter. The result of bit error rate, MSE, and PSNR without raised cosine filter is as shown in Table II. From that result, it has been seen that BER is zero when the E_b/N_0(dB) value is greater than eight and the value of MSC and PSNR is constant for those E_b/N_0(dB) values which is 13.7908 and 7.0874 respectively. The best-quality images are received by the proposed simulation with the E_b/N_0(dB) =5 dB and using the raised cosine filter.



(a) (b) (c) (d) (e)



(f) (g) (h) (i) (j)

Fig. 8. Processing example of output images with different E_b/N_0(dB) value without using raised cosine filter: a) E_b/N_0(dB)=1, b) E_b/N_0(dB)=2,c) E_b/N_0(dB)=3, d) E_b/N_0(dB)=4, e) E_b/N_0(dB)=5, f) E_b/N_0(dB)=6, g) E_b/N_0(dB)=7, h) E_b/N_0(dB)=8, i) E_b/N_0(dB)=9, and j) E_b/N_0(dB)=10

TABLE II. EXPERIMENTAL RESULT OF NO OF ERROR, BER, MSC AND PSNR USING WITHOUT FILTER

E_b/N_0(dB)	Number of error	Bit error rate (BER)	MSC	PSNR
1	29675	0.0116	NA	NA
2	14460	0.0056	NA	NA
3	6121	0.0025	NA	NA
4	2148	$8.3906xe^{-4}$	NA	NA
5	501	$1.9570xe^{-4}$	NA	NA
6	123	$4.8047xe^{-5}$	$8.7112xe^{36}$	$1.1220xe^{-35}$
7	20	$7.8125xe^{-6}$	13.7908	7.0874
8	3	$1.1719xe^{-6}$	13.7908	7.0874
9	0	0	13.7908	7.0874
10	0	0	13.7908	7.0874

Putting the values, for instance, E_b / N_o =1, 2, 3, 4, and 5 dBs respectively, the bit error rate (BER), PSNR is much better than transmitting image without filter. At those values, the retrieve images are blurred which is vague for human visualization using without filter. For the higher values of E_b / N_o, where no error is occurred, conventional communication model with QAM-32 reveals good results. For low signal to noise ratio, the interference caused by inter-symbol interference is dominated over noise. As a result without filter the image is blurred. That suggests that proposed method is more applicable than conventional communication model for image transmission.

The comparison of SNR and BER using with and without filter is as shown in Fig. 9.

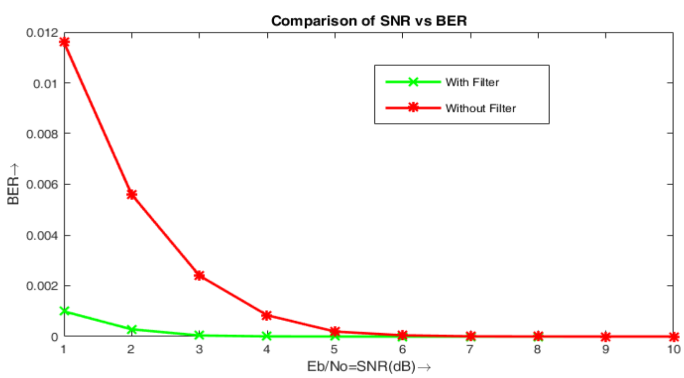


Fig. 9. Comparison of SNR vs BER using with and without filter.

V. CONCLUSIONS

For Low SNR, when compressed image is corrupted by noisy channel, it is hard to retrieve original image. The fact is that signal to interference ratio (SIR) is grater that SNR. For that reason, for lower value of Eb/N0 the images are Vague. However, this problem is mitigated by introducing raised cosine filter in the proposed method. For higher SNR, the conventional communication model with QAM-32 works better. In future, the image quality and data protection may be ensured by implementing Hierarchical QAM (HQAM) and sophisticated coding technique using proposed communication model for image transmission.

REFERENCES

[1] A. Mishra, , K. Sharma, and A. De, "Quality Image Transmission through AWGN Channel using Polar Codes," International Journal of Computer Science and Telecommunication, Vol. 5, No. 1, January 2015.

[2] M. A. Kader,F. Ghani, and R. B. Ahmad, "Image Transmission Over Noisy Wireless Channels Using HQAM and Median Filter," In The 8th International Conference on Robotic, Vision, Signal Processing & Power Applications , pp. 145-152, 2014. Springer Singapore.

[3] Khandokar, M. A. H., Rashed, M. G., Yasmin, R., & Hasan, M. A. (2009). Performance Comparison of M-Ary Modulation Schemes for an Efficient Wireless Communication System with Digital Color Image Transmission.

[4] L. Hanzo, W. Webb, T. Keller, "Basic QAM Techniques" in Single and multi carrier quadrature amplitude modulation: principles and applications for personal communications, WLANs and broadcasting, John W. & S., 2000.

[5] D. K. Chy., M. Khaliluzzaman. "Comparative Performance of BER in the Simulation of Digital Communication Systems using Raised Cosine Filter." Third Intl. Conf. on Advances in Computing, Electronics and Electrical Technology - CEET 2015, doi: 10.15224/ 978-1 -63248-056-9-25, pp-29-33, 2015.

[6] V. Shashank, Maiya, J. C. Daniel and E. F. Thomas, "Low latency Coding; Convolutional Codes versus LDPC", IEEE Transactions on Communications, Vol. 60, No. 5, May 2012.

[7] D. K. Chy., and M. Khaliluzzaman. "Performance Analysis of Digital Communication Systems in Presence of AWGN Implementing Filter Technique," International Journal of Computer Applications, Vol. 131, No. 18, December 2015.

[8] L. Tao and H. K. Kwan, "Multirate-based fast parallel algorithms for DCT-kernel-based real-valued discrete Gabor transform," Signal Processing, vol. 92, no. 3, pp. 679–684, 2012.

[9] A. Sanyal and S. K. Samaddar, "Th performance analysis of fast DCT algorithms on parallel cluster architecture," International Journal of Information and Electronics Engineering, vol. 2, no. 3, pp. 369–373, 2012.

[10] Abhijyoti Ghosh, "Comparative BER Performance of M-ary QAM-OFDM System in AWGN & Multipath Fading Channel", International Journal on Computer Science and Engineering (IJCSE), Vol. 4 No. 06 June 2012.

[11] John G. Proakis and Masoud Salehi, Contemporary Communication Systems using MATLAB. Boston, USA: PWS Publishing Company, 1998.

[12] M. Samsuzzannan ,M. A Rahman and M. A Masud, "Bit Error Rate Performance Analysis on Modulation Techniques of Wideband Code Division Multiple Access," Journal of Telecommunications, vol. 1, no. 2, pp. 22-29, March 2010.

[13] Benard Sklar, Digital Communications: Fundamentals and Applications, 2nd ed. New Jersey, Carlifonia: Prentice Hall.

[14] T. Stathaki, A. G. Constantinides, and G. Stathakis, "Equiripple minimum phase piecewise flat FIR filter design from linear phase systems using root moments," in Proc. IEEE ICASSP 98, Seattle, WA.

[15] K. Z. Bukhari, G. K. Kuzmanov, and S. Vassiliadis, "DCT and IDCT implementations on diffrent FPGA technologies," in *Proceedings of the 13th Annual Workshop on Circuits, Systems and Signal Processing*, pp. 232–235, Computer Engineering Lab, Delf University of Technology, Veldhoven, Th Netherlands, November 2002, http://ce-publications.et.tudelftnl.

Performance Survey of HSPA Network in Chittagong City

Abu Sayed Chowdhury[1], Sadia Kamal[2], Nusera Tasrin[2]
Assistant Professor[1], Student[2], Electronics and Telecommunication Engineering
University of Science and Technology Chittagong
Chittagong, Bangladesh

Abstract— **Demand of mobile and data bandwidth is at highest peak due to huge use of smart phones, laptops, notebooks, and High Speed Packet Access (HSPA) has risen to one of the fastest growing mobile broadband technologies in different markets as well as in Bangladesh. HSPA promised to improve the end user performance and user throughput by lower latency, low power consumption, low cost and increases the peak data rates per user. Popular applications such as web browsing, VoIP, and video streaming enable HSPA to compete with fixed connections. But it is ambiguous that how it performs in live network. This paper investigated the performance of HSPA networks in terms of Received Signal Strength (RxL), Data Throughput, Jitter, Packet Loss Rate, Round Trip Time (RTT), DNS lookup time. The comparison was made based on performance between three different operators in stationary scenarios in Chittagong Metropolitan City.**

Keywords— HSPA, RxL, Throughput, latency, RTT.

I. INTRODUCTION

Mobile cellular system has gone through fundamental changes in recent years. Along with the development of communication technology, the demands are expanding to high speed multimedia services, including Voice-over-IP (VoIP) and video streaming. Many technologies are being introduced to support high-speed data communication. To satisfy the users, operators are introducing High speed packet access (HSPA) technology to improve their 3rd generation/ Universal mobile Telecommunication system (3G/UMTS) network. HSPA technology allows higher data rates, lower delays including the use of real time applications [1] (e.g. VoIP, media streaming, video conferencing, online gaming). As operators are using HSPA to fulfill the consumer's demand, it is important to measure the live network performance of each operator. This measurement can be helpful to understand the network's true behavior [1]. The information provided by the measurement is useful to end users to decide which service provider is more reliable for them.

We measured some parameters of three commercial operators having HSPA-enabled network and analyzed results to understand the difference of performance between them. The three Operators are Grameenphone, Banglalink and Robi Axiata Limited. We measured Received Signal Strength (RxL), data rates through Transfer control protocol (TCP) performance, jitter via User datagram protocol (UDP) performance, Packet Loss Ratio, and latency as Round trip time (RTT) via PING for stationary condition [2].

II. BACKGROUND

HSPA is a combined technology of High speed downlink packet access (HSDPA) and High speed uplink packet access (HSUPA). HSDPA was first introduced in 3rd Generation Partnership Project (3GPP) release 5 in addition to the basic Wideband Code Division Multiple Access (WCDMA). This enhances the downlink (DL) channel performance compared to DL performance of basic WCDMA. The next deployment has came in this pipeline is HSUPA, which was introduced in 3GPP release 6 [3]. In [1], they measured HSPA performance from end user perspective and compared the result to basic WCDMA access. They showed that the first connection throughput phenomenon is still weakening the performance observed by a web user. They also observed the significant affects of mobility to the user experience. As it is stated in [2], they worked on Mobile Network Performance from User Devices considering HTTP GET throughput, round trip time latency from ping, and DNS lookup time. They not only measured the performance of HSPA but also measured the performance of EDGE, UMTS, GPRS etc. they have used two identical apps speedometer and Mobiperf. They found the significant differences in end-to-end performance across carriers, access technologies, and geographic regions and over time. In the paper [4], Kara evaluates the performance of video streaming in HSDPA networks in simple scenarios. Generally mobility degrades HSPA services on the other some aspects of networking performance (e.g. fairness in bandwidth allocation among users, traffic flows) improve because of mobility. Many theoretical studies which are more useful for preliminary capacity approximation and network planning purposes are proposed in prior works. These works focus on HSDPA scheduling or performance evaluation by simulations [5] [6] [7] [8] [9]. However in real 3G and HSPA networks, the theoretical model is hard to formulate, especially in mobile environments. There have also been quite a number of field measurement studies on operational 3G networks, Cano-Garcia [10] and Pentikousis [11] mainly focus on the packet delay behaviour and throughput of pure data traffic under lightly-loaded scenarios respectively. And Tan [12] measures the data throughput and latency of live 3G (WCDMA) networks under saturated conditions, using a mixture of data, video and voice traffic, but which is just in stationary environments under WCDMA. Liu [13] presents an experimental characterization of the physical and MAC layers in CDMA 1xEV-DO and their impact on transport layer performance. Derksen [14] presents the results of HSDPA measurements made in live, commercial networks supplied by Ericsson. But the paper just gives an average downlink throughput in mobile vehicle and not analyzes the factors impact on the performance. Litjens [15] presents a flow level performance evaluation of data transfer in a UMTS/HSDPA network with a principal focus on the performance impact of terminal mobility. But the experiments are carried out in an experimental setup and small scale in a cell. Yao [16] gives an empirical study of bandwidth predictability in the mobile environments. The authors repeat trips along a 23km route in Sydney under typical driving conditions and measure bandwidth from two independent cellular providers implementing the popular HSDPA technology in two different peak access rates (1.8 and 3.6 Mbps). But they only investigate the bandwidth and examine download traffics.

III. HSPA

With some new packages of software and new pieces of hardware in the base station WCDMA can be upgraded to HSPA. Figure 1 shows the HSPA network architecture.

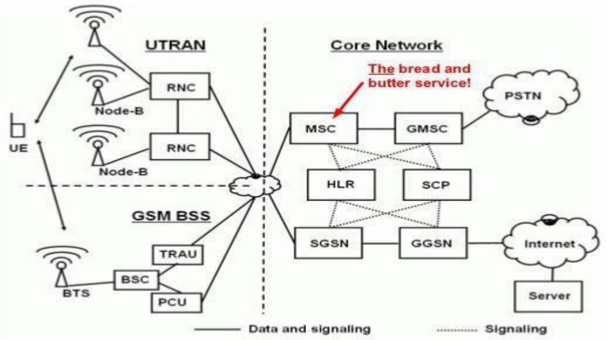


Figure 1. UMTS/HSPA network architecture [17]

The core of HSUPA is the new enhanced dedicated channel (E-DCH).Under ideal radio conditions; HSUPA was theoretically able to deliver uplink speeds up to 5.76 Mbps.

On the other hand Rel.'5 HSDPA is introduced with High Speed Downlink Shared Channel (HS-DSCH), fast HSDPA scheduler, adaptive modulation and coding and fast retransmission (HARQ). Theoretically HSDPA has downlink data rate up to 14.4 Mbps. Beyond the installations of current HSPA, 3GPP release 7 brings further improvements. It introduced enhanced technology of HSPA also called as HSPA evolution or HSPA+. It is able to provide major end-user performance and network efficiency improvements to HSPA. For example, VoIP performance over HSUPA is enhanced by minimizing the control overhead with packet bundling operation by aggregating several VoIP packets into one.

IV. DATA COLLECTION

In Bangladesh, there are three operators (GrameenPhone, Robi and Banglalink) which offer HSPA network service. We chose all of them for survey. We have used three Android operating phones each for one operator. We measure the selected parameters in thirteen different location of Chittagong City for these three operators. These pre-selected places cover important traffic areas in Chittagong city. All our measurements have taken for stationary scenario.

We have used four Apps for this survey. They are free licensed Apps available on Google Play Store. MobiPerf is an open-source application built on the Mobilyzer library. Mobilyzer is a research-oriented measurement library to support client-based measurements.TCP and UDP throughput, RTT, latency measured using this application. G-MON,RF signal tracker and NET-monitor are Android apps using a nearly identical codebase . In every location we run the softwares one by one and took screenshots of measurements. Then we gathered the data from the snapshots.

Received Signal Strength (RxL) is the signal level in dBm received by MS (Mobile Subscriber) from serving Cell. The RxL were measured during the survey by using three different tools G-MoN, NETmonitor and RF Signal Tracker. The data of TCP uplink and downlink speed, Jitter for UDP, PING RTT latency, Packet Loss Ratio are collected by using Mobiperf [18] software.

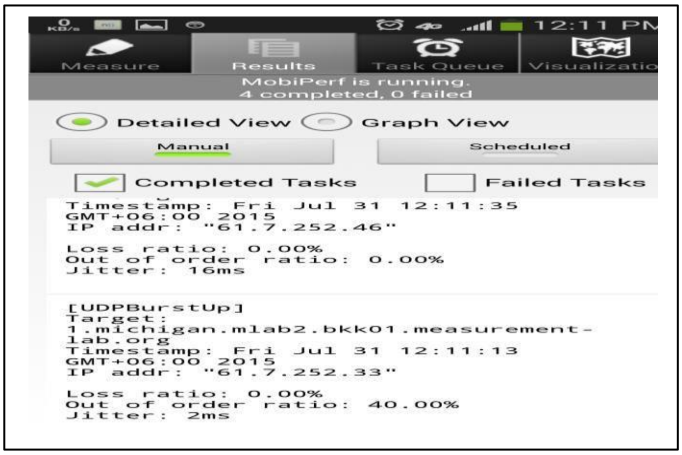


Figure 2. Snapshots of Mobiperf

We got TCP uplink and downlink speed by running the TCP uplink and downlink task respectively. This data actually refers to the throughput using Transmission Control Protocol (TCP). TCP uplink throughput refers to the amount of data received by the Base Transceiver Station (BTS) from MS in per unit time. TCP uplink throughput refers to the amount of

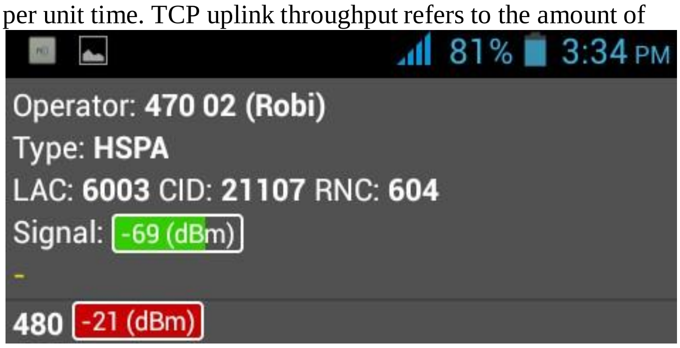


Figure 3. Net Monitor

data received by the MS from BTS in per unit time. Throughput is usually measured in bits per second (bps). Jitter is the time difference in packet inter-arrival time to their destination. Jitter is measured by UDP brust UP and UDP brust

Figure 4. GMON

Down for 10 packets of 10 bytes each. Network latency is a kind of delay. Latency effectively decreases the bandwidth and can be the cause of reducing the overall throughput and Network latency is measured by the packet RTT [19]. RTT

Figure 5. RF Signal Tracker

refers to the time it takes for a signal (packet) to be sent plus the time it takes for an acknowledgment of that packet to be received. We measured RTT by PING. We had PING in Google.com and got RTT along with Packet loss ratio. Packet loss ratio is the percentage of difference between transmitted packets to the received packets. Figure 2, 3, 4 and 5 represent snapshots of four used Apps.

V. RESULTS

Data collected from different places are gathered in graphs. In Figure 6, all the data taken by G-MON is shown, from the graph we can see that highest signal strength is at Chawkbazar and lowest at Rangipara for GP. For Robi, highest peak is at GEC and lowest at Rangipara. For Banglalink, highest peak is at Doublemooring and lowest at Rangipara. Figure 7, shows the signal strength taken by NETmonitor at different places. Result is same as G-MON for all operator. Figure 5, exhibits data of RF signal tracker which also similar with previous. These figures pointed that the best received signal strength is in GP. Robi and BL are almost incident on each other. At Rangipara all operator signal is poor.

Figure 9 and Figure 10 show the TCP throughput at different places. It is seen that throughput is less in GP and high in BL. Figure 11 and Figure 12 shows the jitter performance at survey areas for UDP uplink and downlink respectively.

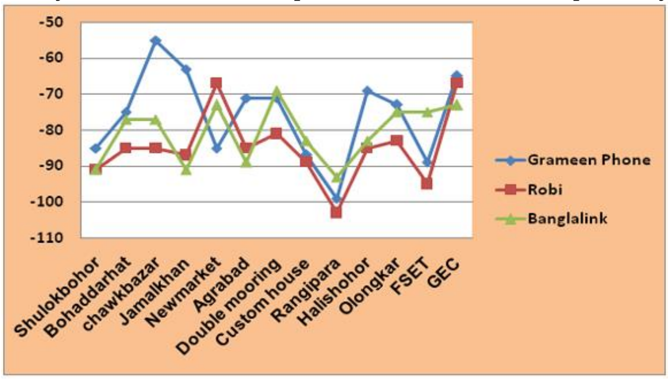


Figure 6. RxL(dBm) at different places using G-Mon

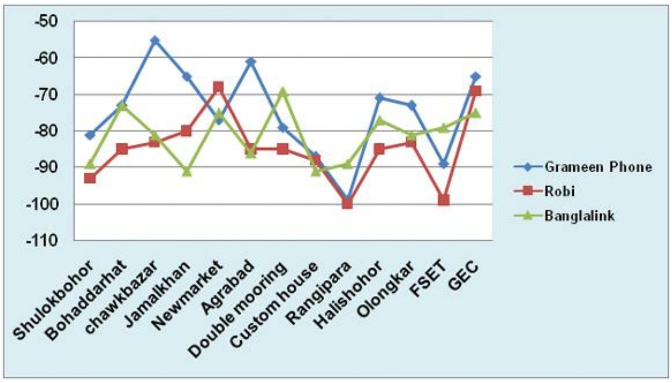


Figure 7. RxL(dBm) at different places using NETmonitor

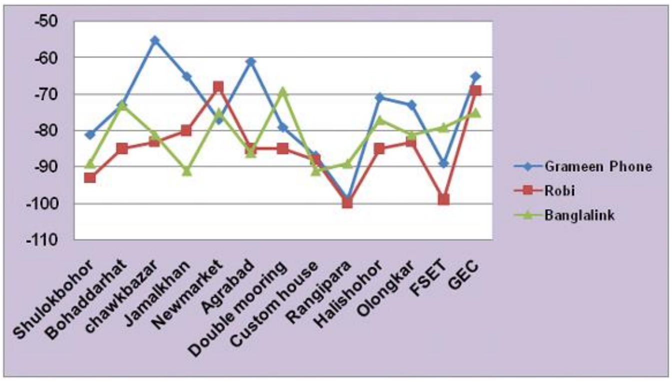


Figure 8. RxL(dBm) at different places using RF signal tracker

Figure 13 shows the latency measured as RTT. It is the higher latency in GP. Figure 14 shows the packet loss ratio which is higher in GP. Table I depicts the overall comparison among three operators.

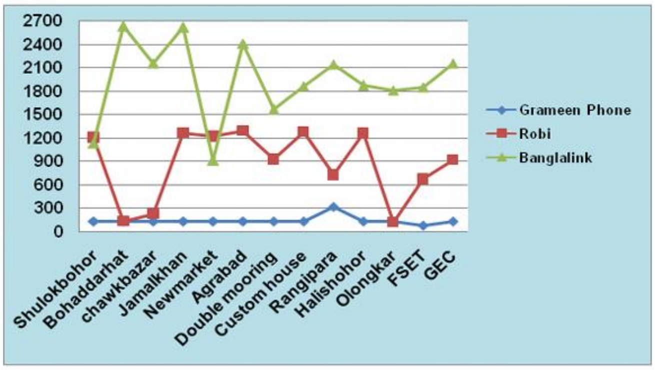


Figure 9. TCP uplink speed (kbps) at different places

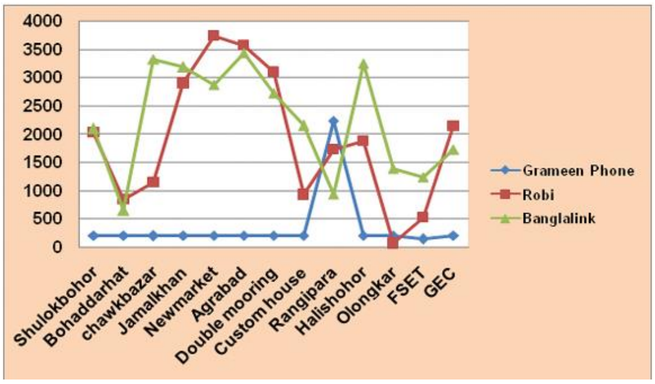


Figure 10. TCP Downlink speed (kbps) at different places

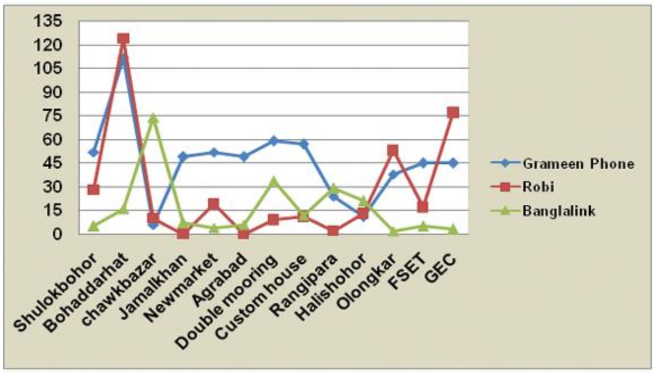


Figure11. UDP uplink Jitter (ms) at different places

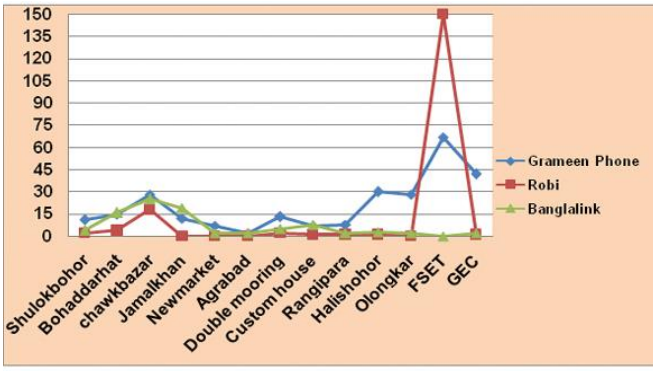


Figure 12. UDP downlink Jitter (ms) at different places

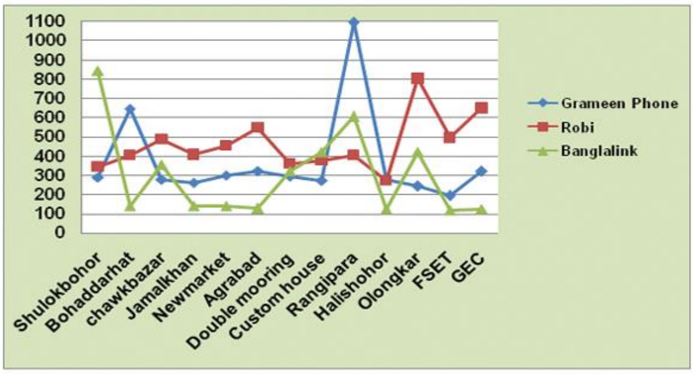


Figure 13. Mean RTT (ms) at different places

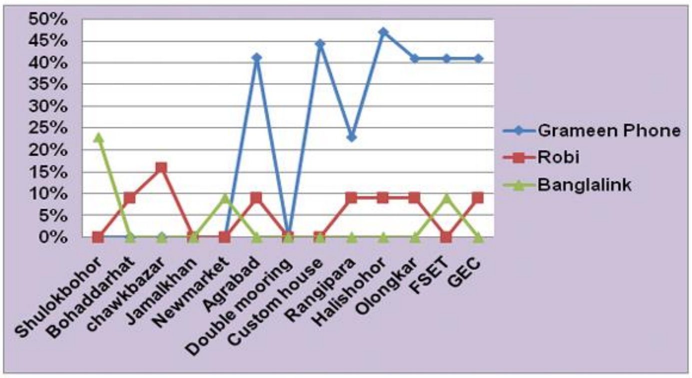


Figure 14. Packet Loss Ratio at different places

It is clearly observed that Grameenphone prefers to do business with voice call since it has greater signal strength. On the other hand Robi and Banglalink are focusing on data since TCP and UDP parameters providing better results.

Table I. Comparing Performance of Three Operators

	Best	Worst
Signal strength	Grameen Phone	Robi
TCP Uplink Speed	Banglalink	GP
TCP Downlink Speed	Banglalink	GP
Jitter for UDP Burst Up	Banglalink	GP
Jitter for UDP Burst Down	Robi	GP
Packet Loss Ratio	Banglalink	GP
Mean RTT	Banglalink	Robi
DNS Lookup Time	Robi	Banglalink

From above figures and tables it is seen that GP provides best RxL. But it failed to provide proper throughput due to having a plenty of cell with overlapping area more than ideal one (20%). That causes neighbor cell interference which estimates cell capacity and drive to give poor throughput. Where else other two operators have comparatively lower RXL, but provide better throughput. Robi and Banglalink have less number of BTS and less overlapping area. Jitter is maximum in GP that causes voice quality damped; packet loss is result of jitter which is obviously maximum in GP. During our survey we did not get a single jitter time within threshold for GP. Here Banglalink provides a suitable and efficient jitter thus hinder the packet loss. RTT is the principal matrix of quality measurement. In networks with packet acknowledging schemes such as TCP/IP, the maximum effective data throughput is not necessarily equal to the system's peak rate: the latency can reduce the overall throughput due to the time required to acknowledge the data packets. The fact that all packets being transmitted wait for an acknowledgment by the peer entity requires the presence of a packet queue on the transmitter (also called transmission window where packets will sit waiting until they have been acknowledged by the receiver. Since there are physical or practical limitations to the size of the packets being transmitted (network Maximum Transfer Unit or MTU), and to the size of the transmission window, if the packet acknowledgment time – which is directly related to the system's latency – is too long, this queue will soon fill up and no more new packets will be able to be transmitted, thus effectively limiting the overall throughput. So, it may well be that your physical transmission media allows a very high speed, however, the practical speed will be reduced to a fraction of that bit rate if the network latency (measured by the packet Round Trip Time) is too high. RTT must be within -70dBm. But no operator in Chittagong able limits RTT within threshold. Maximum is in GP. So, subscriber experience low quality real time applications like web browsing and VoIP.

VI. Conclusion

This report analyzed the end-to-end performance as seen from mobile devices. Measuring parameters such as DNS lookups and RTT, jitter, we identified the reasons behind performance problems. Overall, we find that performance of HSPA is not improving in Bangladesh.

- General concept, the higher the received signal strength, network is better performer. But it is not true, from our survey data and analysis we saw that RTT is the main quality parameter of HSPA, as real time applications are prime importance of this network. RTT decrease the system throughput significantly.
- We propose GP to re-configure core network. It is important to note that reductions in end-to-end delay could come from different parts of the network. For instance, a good planning and distribution of the GPRS core nodes can be very helpful in reducing network latency, especially for HSPA networks in which RAN delay is view, it is desirable to have a distributed SGSN/GGSN network rather than concentrating the min gigantic packet switching centers.
- Banglalink should plan for more BTS with proper coverage area in order to provide good signal strength throughout the whole PLMNA. As Banglalink provides comparatively better TCP performance, thus this measure will satisfy subscriber with better experience.
- Robi shows an average performance in both cases mentioned above. So this operator needs both measures.

At the end, based on promises of HSPA we can say that no operator provides proper services in Chittagong. So, we need troubleshoot and proper planning of HSPA network technology in order to experience the promised features properly.

References

[1] JarmoProkkola, Pekka H. J. Perälä,MikkoHanski, and EsaPiri "3G/HSPA Performance in Live Networks from the End User Perspective" , IEEE ICC 2009

[2] Ashkan Nikravesh1, David R. Choffnes2, Ethan Katz-Bassett3, Z. Morley *Mao1'' Mobile Network Performance from User Devices*: A Longitudinal, Multidimensional Analysis",2013.

[3] H. Holma, et al., "High-Speed Packet Access Evolution in 3GPP Release 7," In IEEE Comm. Magazine, vol 45, Dec 2007, pp 29-35.

[4] N. Kara and V. Planat. Performance analysis of IP multimedia services over HSDPA mobile networks. pages 1–5, Dec. 2007.

[5] K. Aho, J. Aijanen, and T. Ristaniemi. "Impact of mobility to the VoIP over HSUPA system level performance. pages 2091–2095, May 2008.

[6] M. L´opez-Ben´ıtez, F. Bernardo, N. Vucevic, and A. Umbert. "Real-time HSPA emulator for end-to-edge QoS evaluation in all-ip beyond 3G heterogeneous wireless networks," In SimuTools, page 8, 2008.

[7] P. Lunden, J. Aijanen, K. Aho, and T. Ristaniemi. Performance of VoIP over HSDPA in mobility scenarios. pages 2046–2050, May 2008.

[8] V. Vukadinovic and G. Karlsson. Video streaming in 3.5G: On throughput-delay performance of proportional fair scheduling. pages 393–400, Sept. 2006.

[9] X. Yan, J. Khan, and B. Jones. Impact of a radio access network capacity on the HSDPA link performance. pages 220–224, 30 2007-Oct. 3 2007.

[10] J. M. Cano-Garcia, E. Gonzalez-Parada, and E. Casilari. Experimental analysis and characterization of packet delay in UMTS networks. volume 4003 of Lecture Notes in Computer Science, pages 396–407. Springer, 2006.

[11] K. Pentikousis, M. Palola, M. Jurvansuu, and P. Perala. Active goodput measurements from a public 3G/UMTS network. Communications Letters, IEEE, 9(9):802–804, Sep 2005.

[12] W. L. Tan, F. Lam, and W. C. Lau. An empirical study on the capacity and performance of 3G networks. IEEE Transactions on Mobile Computing, 7(6):737–750, June 2008.

[13] X. Liu, A. Sridharan, S. Machiraju, M. Seshadri, and
H. Zang. Experiences in a 3G network: interplay between the wireless channel and applications. In MOBICOM, pages 211–222, 2008.

[14] J. Derksen, R. Jansen, M. Maijala, and E. Westerberg. HSDPA performance and evolution. Ericsson Review, (3):117–120, 2006.

[15] R. Litjens. HSDPA flow level performance and the impact of terminal mobility. volume 3, pages 1657–1663 Vol. 3, March 2005.

[16] J. Yao, S. S. Kanhere, and M. Hassan. An empirical study of bandwidth predictability in mobile computing. In WINTECH, pages 11–18, 2008.

[17]Slideshare,http://www.slideshare.net/wirelessmoves/from-gsm-to-lte last viewed 25/08/15.

[18] Mobiperf, https://sites.google.com/site/mobiperfdev/,last viewed 25/08/15.

[19] Pablo Tapia, Jun Liu, Yasmin Karimli, T-Mobile USA, Martin J. Feuerstein, Polaris Wireless, USA, "HSPA performance and Evolution," John Wiley & Sons Ltd, 2009.

Big Data Characteristics, Value Chain and Challenges

Rabiul Islam Jony[1], Rakibul Islam Rony[2], Abiduzzaman Rahat[1], Musfiqur Rahman[1]
[1]University of Liberal Arts Bangladesh
{rabiul.islam, abiduzzaman.rahat.ete, musfiqur.rahman.ete}@ulab.edu.bd
[2]Primeasia University
iut.rony@gmail.com

***Abstract*—Recently the world is experiencing an deluge of data from different domains such as telecom, healthcare and supply chain systems. This growth of data has led to an explosion, coining the term Big Data. In addition to the growth in volume, Big Data also exhibits other unique characteristics, such as velocity and variety. This large volume, rapidly increasing and verities of data is becoming the key basis of completion, underpinning new waves of productivity growth, innovation and customer surplus. Big Data is about to offer tremendous insight to the organizations, but the traditional data analysis architecture is not capable to handle Big Data. Therefore, it calls for a sophisticated value chain and proper analytics to unearth the opportunity it holds. This research identifies the characteristics of Big Data and presents a sophisticated Big Data value chain as finding of this research. It also describes the typical challenges of Big Data, which are required to be solved. As a part of this research twenty experts from different industries and academies of Finland were interviewed.**

***Key words*—Big Data, Big Data characteristics, Big Data Value chain, Big Data Challenges.**

I. INTRODUCTION

In the year 2000, when the Sloan Digital Sky survey started their work, its telescope in New Mexico collected more data on its first few weeks than had been amassed in the entire history of astronomy. After one decade, now its archive contains around 140 terabytes of data. Another large synoptic Survey Telescope in Chile is predicted to collect the same quantity of data every five days by 2016 [1]. Wal-Mart, the retail giant, generates around 2.5 petabytes data of 1 million customers' transactions every hour [2]. Facebook, a social networking website stores 500+ terabytes of new data every day. Search engines like Google processes 20 petabytes of data every day [1]. All these examples show how much data the world contains and how rapidly the volume of data is growing. Until 2003, 5 exabytes of data were created by humans, and currently this amount of data is created in only two days [3]. The amount of data in the digital world reached 2.72 zettabytes in 2012 and is expected to double in every two years reaching 8 zettabytes by 2015 [4].Data is getting so large and complex, that it is becoming difficult to process using traditional data processing applications, and introducing big data.

The most popular definition of big data is defined by Gartner as "Big data is high-volume, high-velocity and/or high-variety information assets that require new forms of processing to enable enhanced decision making, insight discovery and process optimization" [5].

Big data is currently treated as a technology, which has been developed to handle large volumes of fast-changing and non-schematic data. Big data technology also provides companies, such as telecom operators, with an ideal platform for centralizing and storing and analyzing their structured, unstructured and semi-structured data. These yield major advantages in data analysis, knowledge discovery and new business opportunity identification.

According to McKinsey Global Institute (MGI) research, big data is the key basis of competition, underpinning new waves of productivity growth, innovation and customer surplus of the future market [6]. Therefore, organizations need to have a clear idea about the characteristics and value chain of Big Data. Big Data also comes with potential challenges, which needs to be attended well before starting Big Data initiatives.

In this paper, section II describes the characteristics of Big Data. Section III describes the value chain of Big Data, which is a finding of this research. In section IV the potential Big Data challenges are presented. Finally, the paper is concluded in section IV.

II. BIG DATA CHARACTERISTICS

The characteristics of big data are well defined in the definition by Gartner. The three Vs (volume, velocity and variety) can be considered as the main characteristics of big data. These characteristics are described below.

A. Volume

Data volume measures the amount of data available to an organization; the organization does not necessarily have to own all of it as long as it can access it [7]. The number of sources of data for an organization is growing. More data sources consisting large datasets increase the volume of data, which needs to be analyzed. As data volume increases, the value of different data records decreases in portion to age, type, richness and quality among the other factors [7].

Figure 1 indicates the volume of data stored in the world by year. It also predicts that the amount of data would be more than 40 zettabytes (10^{21}) by 2020 [8].

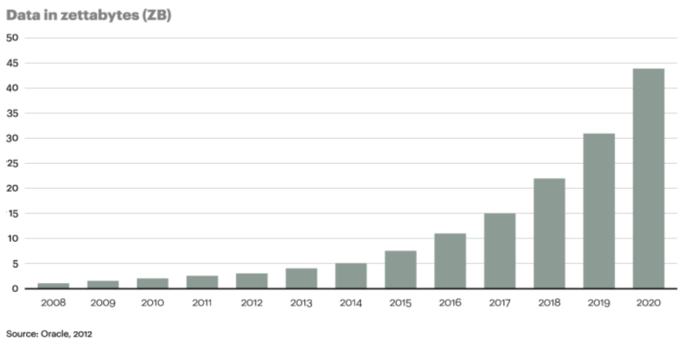


Figure 1: Data volume growth by year in zettabytes [8]

B. Velocity

Data velocity measures the speed of data creation, streaming and aggregation [7]. According to Svetlana Sicular from Gartner, velocity is the most misunderstood big data characteristic [9]. She describes that the data velocity is also about the rate changes, and about combining data sets that are coming with different speeds. She also argued that, the velocity of data also describes bursts of activities, rather than the usual steady tempo where velocity frequently equated to only real-time analytics [9].

C. Variety

Other than typical structured data, big data contains text, audio, images, videos, and many more unstructured and semi-structured data, which are available in many analog and digital formats. From analytics perspective, variety of data is the biggest challenge to effectively use it. Some researchers believe that, taming the data variety and volatility is the key of big data [2]. Data variety is also considered as a measure of the richness of the data presentation. Incomputable data formats, non-aligned data structures and inconsistent data semantics represents significant challenges that can lead to analytic sprawl [7].

Figure 2 shows the comparison between increment of unstructured, semi-structured data and structured data by years, which reflects the increment in variety of data.

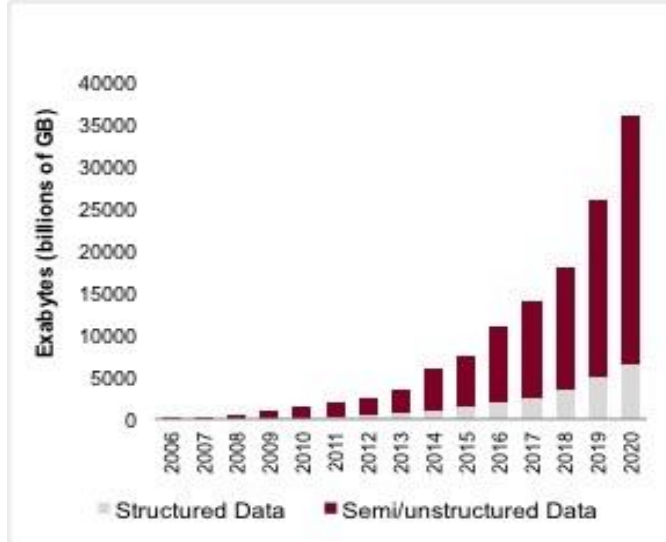


Figure 2: Growth of data variety by year

One of the big data vendors, IBM has coined additional V as the big data characteristics, which is veracity. By veracity, they address the inherent trustworthiness of the data. As big data will be used e.g. for decision making, it is important to make sure that the data can be trusted.

Some researchers mentioned 'viability' and 'value' as the fourth and the fifth characteristics leaving 'veracity' out [10]. The characteristics of big data can also be described with HACE theorem. The theorem states that, "Big data starts with large-volume; **h**eterogeneous, **a**utonomous sources with distributed and decentralized control and seeks to explore **c**omplex and **e**volving relationships among data [11]. From the theorem the key characteristics are defined as:

1) Huge Data with Heterogeneous and Diverse Dimensionality: Here the 'heterogeneous' feature refers to the different types of representations for the same individuals. The feature 'diverse' reflects the variety of the features involved to represent each single observation.

2) Autonomous Sources with Distributed and Decentralized Control: 'Autonomous' feature describes the ability of each data sources to generate and collect information without any centralized control.

3) Complex and Evolving Relationships: With volume of data the complexity and the correlations among them increases.

In summary, based on the characteristics described above, big data can be defined as large volume, high velocity and verities of data, which is complex to process with traditional applications, but able to bring new business opportunities to the industries by enhanced insight generation

III. Big Data Value Chain

Few decades ago, Michale E. Porter first introduced the concept of value chain, where he explained a value chain as a series of activities that create and build value as it progresses [12]. Finally, these activities culminated in total value, which the organizations then deliver to its customer [13]. In 1988, R. L. Ackoff first specified data value chain [14], which was a hierarchy based on filtration, reduction, and transformation, showing how data lead to information, knowledge and finally to wisdom. He presented Data-Information-Knowledge-Wisdom hierarchy as a pyramid which produces a series of opposing terms including misinformation, error, ignorance and stupidity when inverted [15]. Ackoff has fitted wisdom on the top of the hierarchy pyramid followed by knowledge, information and then the data or the raw data.

Figure 3: The Data-Information-Knowledge-Wisdom hierarchy pyramid

Table 1 below describes the four components of Data-Information-Knowledge-Wisdom hierarchy

Category	Description
Data	Data is raw. It simply exists and has no significance beyond its existence and it does not have any meaning of itself. Data can also be defined as Computerized representation of models and attributes of real or simulated entities.
Information	Information is the data that has been given meaning by way of relational connection. Information can also be defined as the data that represents the results of the computational process such as statistical analysis, providing answers to questions, such as 'who', 'what', 'where', and 'when'.
Knowledge	Knowledge is the appropriate collection of the information calculated out of raw data and its intent has to be useful. Knowledge might also be defined as the data that represents the results of a computer-simulated cognitive process, such as perception, learning, and reasoning. Knowledge is the application of data and information which provides the answers to 'how' questions.
Wisdom	Wisdom represents the ability to see the long-term consequences of any act and evaluate them relatively to the ideal of total control. Wisdom is a non-deterministic and non-probabilistic process that answers questions like 'what needs to be done and why'. Wisdom can also be defined as the process by which the outcome can be judged.

Table 1 Data-Information-Knowledge-Wisdom hierarchy

The Big Data value chain in this research is divided into three stages, naming Data sources, Preprocessing and storing, and Processing and Visualization, where each step increases value.

A. *Data sources, types and accessibility*

Data source is the first stage of the Big Data value chain. The data types and accessibility are also included in the sources tag because these also define the value. This step can be divided into three sub-divisions naming availability, amount and accessibility. These define the value of the data sources.

In Figure 4 below the Big Data value chain is presented. Where we can see that, difficult to easy is mentioned in this step, which means if the data from the sources is easily accessible, it has higher value.

This stage of the value chain lies under the data section of the Data-Information-Knowledge-Wisdom pyramid. For ease of graphics design, the pyramid is drawn horizontally in the value chain.

B. *Preprocessing and Storing*

This stage of the value chain brings the information out of the data, and belongs to the information part of the pyramid.

Value increases with the capability of colleting, loading, and preparing the data. There are different kinds of data types in big data, and capability of reading all types of data increases value. The data preparing capability also increases the value. Typically data needs to be stored in this phase, but if real-time analysis is required, data might be stored after the actual analysis.

The preprocessing step of the value chain reflects the ETL (Extract-Transform-Load) process. The ETL process is described below.

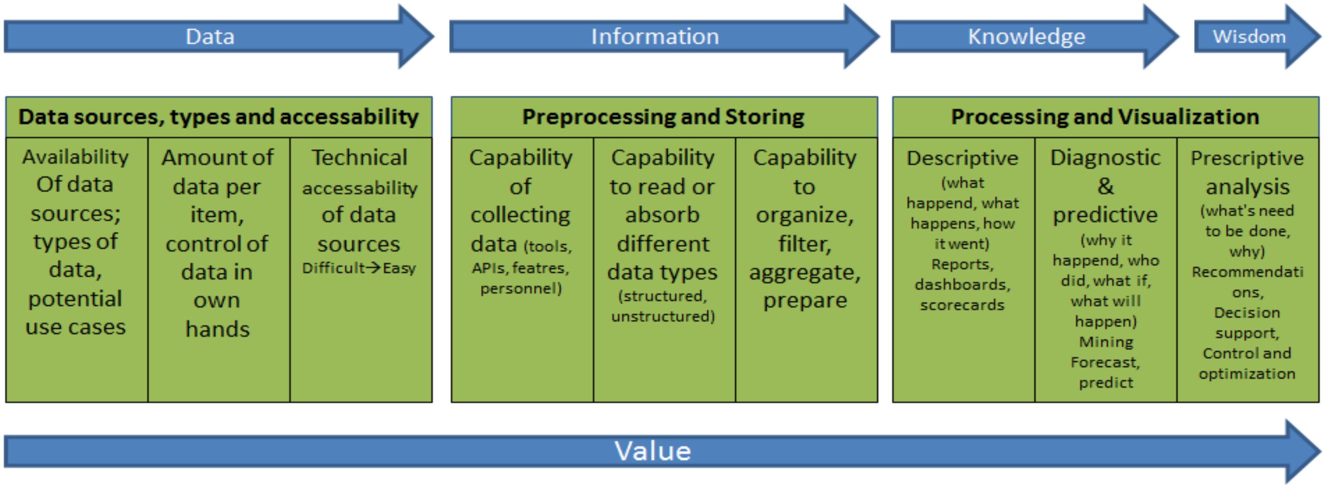


Figure 4: Big Data Value Chain

- **Extract** is the first step of ETL process which covers the data extraction from the source system and makes it accessible for further processing. The goal of this step is to retrieve required data from all the sources with little resources, and not to affect the process in terms of performance, response time negatively. Data extraction can be performed in several ways like update notification, incremental extraction and full extraction.
- **Transform** step cleans the data, which is important to ensure the quality of the data. When the data is cleaned then transform step applies a set or rules to transform the data from source to target. This includes several tasks, such as translating coded values, encoding free-from values, sorting, joining the data from multiple sources, aggregation and splitting according to the application requirements.

- **Load** phase loads the transformed data into the end target. Depending on the requirements of the applications, this process varies widely. Typically the target of the load phase is the databases or data warehouses. During the load step it is also necessary to ensure that the load is performed correctly with the minimal resources usage.

The ETL process framework is shown in the Figure 5 below.

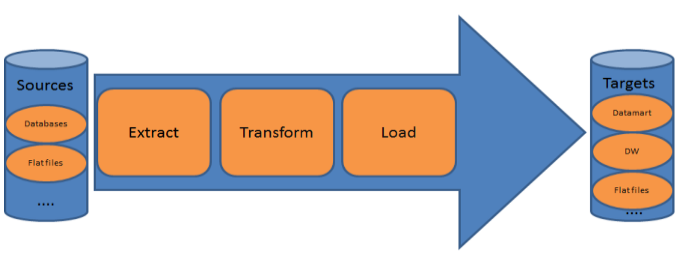


Figure 5: Typical ETL process framework

C. *Processing and Visualization*

This stage of the value chain creates the highest value and can also be called as 'Analytics and Visualization'. It lays into both knowledge and wisdom parts of the pyramid. Descriptive analysis works on past and present results and answers questions, such as what happened, what happens, and how it went. On the other hand, diagnostic and predictive analysis investigate the results and answer questions, such as why it happened, who did and what is going to happen. Both the processes increase value and bring knowledge. Prescriptive analysis works on future and includes questions, such as what is needed to be done and why, also brings wisdom. Wisdom has the highest value in the value chain [16].

The big data value chain shows the process of converting data into information, knowledge, and finally to wisdom. It also shows that every stage plays an important mediator role to enable information and generates insights from data.

IV. Big Data Challenges

Big data also has some significant challenges, some of them are mentioned below:

- **Storage**: The first and foremost challenge of big data is the storing. Traditional data warehouses are not typically made for it. Organizations that are trying to adopt big data strategy need to build a new warehouse, which is capable of storing big data for them.

- **Complexity**: The three dimensions of big data, namely volume, velocity, and variety make it more complex and challenging to analyze than the other traditional data.

- **Management**: Big data management systems available in the current market are not able to satisfy the needs of it, thus a re-construction of the information framework is needed. Re-organizing the data in this re-constructed frame work is another big challenge.

- **Preprocessing**: Finding out the right data from big amount of data which also have verities in it, is typically challenging. Big data preprocessing requires collection capability, statistical analysis, and integration capability of large amount of data. Traditional extract-transform- load (ETL) tools are not able to fulfill these requirements.

- **Analytics**: Big data analytic is a highly mathematics intensive analytic modeling exercise which requires proper tools and skilled people. Big data also requires highly capable tools for data visualization, because traditional tools are typically made for small amount of data.

- **Utilization gap**: Christine Moorman stated that the biggest challenge regarding big data is the Utilization gap [17]. When asked to report the percentage of project in which their companies are using marketing analytics that are available, CMOs report a dismal of only 30% usage rate [17]. In [18], the hardest challenge of big data is mentioned as, taking the insights generated from the analytics and utilizing them to change the way business operates.

- **Lack of skilled people**: As big data is a new concept and requires newer technologies; there is a lack of skilled people for it. According to Gartner, big data demand will reach around 4.4 million jobs globally by 2015, with two third of these positions remaining unfilled [19].

- **Privacy**: In several research studies, privacy concern has defined as the biggest barrier for big data [7], [20]. When it comes to the customer personal data and how it is used, people generally don't like surprises. The study [20] shows, how the location data and the social media data are hampering users' privacy, and the users are not concerned about it. The social media data is being one big topic about the users' privacy issue in recent days but the user location data being as a privacy issue has not got that much attention yet [20].

- **Security**: Big data security management is also one challenging task. Traditional security mechanisms, which are tailored to secure the small-scale data, are inadequate for it.

- **Real-time analysis**: Big data requires real-time analysis, which is sometimes challenging. Real-time analysis requires high-velocity streaming analysis of big amount of data and typical data analysis tools are incapable of doing so.

- **Additional challenges**: There are more additional challenges regarding big data, such as transportation of data, dynamic design requirement, and scaling.

An organization, before starting big data projects needs to make new policies to mitigate these challenges, and select tools which are truly capable for it. Otherwise, the project acquires a large possibility to be failed in the middle of the process which will cause the organization financial loss.

V. Conclusion

In this paper the basic characteristics of Big Data is described and then a summarized definition of Big Data is presented. This paper also presents a Big Data value chain as a result of this research. The value chain defines how the value of data increase as the process continues. The challenges of Big Data this paper presents are based on internet research and literature study. The challenges may also vary from organization to organization depending on their resources and data sizes. As the field of Big Data is still in a great state of change and development, the possibilities for research in this area are extensive and the interest on the findings will certainly rise if the adoption of big data rises as expected. The validation of the result could not be presented in this paper, which requires further research.

VI. Bibliography

[1] The Economist, "Data, data everywhere," *The Economist*, February 2010.

[2] Infosys, "Big Data: Challenges and Opportunities," 2013.

[3] Eric Schmidt. (2010, August) techcrunch. [Online]. http://techcrunch.com/2010/08/04/schmidt-data/

[4] S. Sagiroglu and D. Sinanc, "Big data: A review," in *Collaboration Technologies and Systems (CTS), 2013 International Conference on*, San Diego, CA, 2013, pp. 42 - 47.

[5] Mark A. Beyer and Douglas Laney, "The Importance of 'Big Data': A Definition," Gartner, Analysis Report G00235055, 2012.

[6] McKinsey & Company, "Big Data: The next frontier for innovation, competition, and productivity," 2011.

[7] S. Kaisler, F. Armour, J. A. Espinosa, and W. Money, "Big Data: Issues and Challenges Moving Forward," in *System Sciences (HICSS), 2013 46th Hawaii International Conference on*, Wailea, Maui, HI, 2013, pp. 995-1004.

[8] ATKearney. (2013, January) Big Data and the Creative Destruction of Today's Busniness Model. [Online]. https://www.atkearney.com/strategic-it/ideas-insights/article/-/asset_publisher/LCcgOeS4t85g/content/big-data-and-the-creative-destruction-of-today-s-business-models/10192

[9] S. Sicular, "Gartner's Big Data definition consists of three

parts," *Forbes*, March 2013.

[10] Neil Biehn. (2013) WIRED. [Online]. http://www.wired.com/insights/2013/05/the-missing-vs-in-big-data-viability-and-value/

[11] X. Wu, X. Zhu, G.Q. Wu, and W. Ding, "Data Mining with Big Data. Knowledge and Data Engineering," *Knowledge and Data Engineering, IEEE Transactions on*, vol. 26, no. 1, pp. 97-107, June 2013.

[12] M.E. Porter, *The Competitive Advantage: Creating and Sustaining Superior Performance*. New tork, USA: Free press, 1985.

[13] H. Miller and P. Mork, "From Data to Decisions: A Value Chain for Big Data," *IT Professional*, vol. 15, no. 1, pp. 57-59, February 2013.

[14] R. Ackloff, "From Data to Wisdom," *Journal of Applied Systems Analysis*, vol. 16, pp. 3-9, 1989.

[15] Jay H. Bernstein, "The Data-Information-Knowledge-Wisdom Hierarchy and its Antithesis," *North American Symposium on Knowledge Organization*, pp. 65-75, 2011.

[16] Andrew Stein, "Big Data and Analytics, The Analytics Value Chain – Part 3," *Steinvox*, October 2012.

[17] C. Moorman, "The Utilization Gap: Big Data's Biggest Challenge," *Forbes*, March 2013.

[18] McGuire. (2013, March) Youtube. [Online]. https://www.youtube.com/watch?v=Sc5FFY-IVDQ

[19] Gartner, "Gartner Reveals Top Predictions for IT Organizations and Users for 2013 and Beyond," 2012.

[20] M. Smith, C. Szongott, B. Henne, and G. Von Voigt, "Big data privacy issues in public social media," in *Digital Ecosystems Technologies (DEST), 2012 6th IEEE International Conference on*, Campione d'Italia, 2012, pp. 1-6.

Automatic face annotation with face prototype map in personal photo frame application

Supratip Ghose
Dept. of Computer Science & Engineering
University of Information Technology & Sciences
Baridhara, Dhaka, Bangladesh
SGresearch@gmail.com

Ziku Dhar
Dept.of IT
University of Information Technology & Sciences
Baridhara, Dhaka, Bangladesh
dharma.bead@gmail.com

***Abstract*— In recent years, there is growing trends to form a personal space with the social community by keeping and tagging photos. In this research, we have developed a particular photo frame system, namely, PFS. PFS maintains individual windows of photos and support photo search by drawing the visual content on Canvas. We employed face detection and recognition in image annotation to annotate and search pictures in a flexible manner. The paper proposes a fusion scheme of the classifier for auto-annotation regarding face annotation. We created a framework to find a prototype facial map with the adopted fusion scheme for the classifier. Experimental evaluation of classifier with the test images returns a reasonable hit for image classification by overcoming the false region detection. We show the effect of the quantity of training data on the accuracy of the predictions. (*Abstract*)**

Keywords—Personal Photo Frame; Image Annotation; image processing

I. INTRODUCTION (*HEADING 1*)

In recent years, increasing presence of a user in space like facebook and twitter drives people to maintain personal photos in digital spaces. Web 2.0 user in the social space finds growing resource to share their experience and keep all the profile for their memory. Moreover, many research projects and commercial products have deployed the IT evolution of processing powers and modern sensing devices, analysis and rendering equipment and technologies. The widespread use of these phenomena has created a strong urge for digital content annotation techniques [4] in applications like Facebook and Flickr. Given an arbitrary image, the face detection determines all image regions in a picture that contain a face and returns extents of each face in location and sizes given that it exists. The face detection is the prima facie to understand the desires and intentions of people by facial analysis algorithm. This automatic recognition is one of the challenging tasks with the object class detection in which, the task is to search the locations and sizes of varieties of objects in digital images pertain to a given category. Image auto-annotation research pursues some techniques that find a correlation between low-level visual features and high-level semantics [6]. These face annotation functionalities ensure efficient and effective search in personal photo collection in photo sharing sites in which the uploaded images invites automatic annotation. The search based face annotation paradigm aims to tackle the automated face recognition task by exploiting content-based image retrieval [5]. Noticeable works reported in the literature, over the years, concerned, among other research of training based classifiers for annotation scheme like Support Vector Machine [12], Artificial Neural Network [3] and Decision Tree (DT) [11]. Compared to other learning methods, DT is simple to interpret and understand and can learn with a small number of samples. It is also robust from incomplete and noisy data. The Haar-like feature has been popularly employed for object representation. Papageorgiou et al. [14] proposed to represent objects in terms of a subset of an over-complete dictionary of Haar wavelet basis functions, and Viola and Jones [1] developed the Haar-like feature by extending the idea of using Haar wavelets. Viola and Jones introduced fast object detector based on the Haar-like feature. A new image representation called integral image was introduced to compute Haar-like features efficiently, and AdaBoost [3,4] was used both to select a small number of most discriminating features and to train classifiers. Classifiers were combined in a cascade which allows background regions to be quickly discarded while spending more computation on object-like regions [1],[2].

Recently, some efforts aside commercial perspective, capture user's search intention. Those visually allow them to describe the picture content and layout on a particular query canvas [4], [6], [7]. To the aspect of image similarity, intensity-based metrics and geometry-based metrics [1] are the two forms of classification for objective based image similarity. Intensity-based similarity metrics assumes that the registered images are comparable at the same scale, a comparison of the corresponding pixel intensities determines their similarity. As for this category of metrics, the most commonly used metric is Euclidean distance, which converts images into pixel by point operations. However, neighboring image pixels are highly correlated with each other. To this advantage, geometry-based metrics includes geometric transformation between corresponding pixels based on intensity between the pictures that we adopted in our system. In our system, we provide the annotation facility in a flexible manner estimating face attributes and face similarity. The contribution of the work is that we integrate the annotation scheme in personal Photo framework (PFS) with a fusion scheme that will ultimately help the system forming auto-annotation. We form the notion of visual saliency region compared to the research of Facial landmark Detectors [15]. Facial landmark detectors can be categorized into global and local detectors. Global facial landmark detectors rely on global statistical relations between landmarks, but do not sufficiently utilize local appearance information, whereas local detectors mainly focus on local appearance attributes of landmarks.

In turn, it will overcome the time-intensive manual annotation. The group annotation in the current photo album system still requires making one or more decision for each picture. Moreover, images in the current photo frame system exist in different hierarchies that are difficult to search. Having justified the feasibility of our work, the rest of the paper appears as follows. We describe the system overview in section 2 in which discusses on forming prototype facial map by fusion scheme for auto-annotation and functionalities of the system. In section 3, we carried out the experimentation for creating auto-annotation and provide the effectiveness of face annotation. Next, in Section 4, we draw the conclusion and explore the future work with the system

I. System Overview

In this paper, we have deployed a platform for Photo Frame system (PFS), that allow the user to manipulate and search photo based on the facial appearance. The aim of this research is to help users maintain a personal window of photos and to design and develop a system that conforms user-friendly user experience maintaining usability. We solicit here to make an application based on the existing annotation algorithms in the automatic and manual form. The efficient face annotation scheme recognizes the faces and annotates them properly. An implementation of a canvas ensured graphical manipulations like draw, type, scale, and drug for interactive searches. Our system consists of Browse panel to browse the windows; Search Panel to search photos and Mode Panel to manage people appeared in a photo in Manual and auto-annotation mode. We have integrated face detection and recognition technologies and facilitates user to experience the exploration by providing rich, user-friendly experience for search and query.

A. Facial Color Modelling

The first revolves around the indexing of pixel intensities relative to the current estimate of the shape. The extracted features in the vector representation of a face image can greatly vary due to both shape deformation and nuisance factors such as changes in illumination conditions. This makes accurate shape estimation using these features difficult. The dilemma is that we need reliable features to accurately predict the shape, and on the other hand we need an accurate estimate of the shape to extract reliable features.

To model perceived image similarity, in this section we describe the prototype facial model by color. In the face detection using color, the prior probability of a pixel belonging to a face can be estimated by it feature distance to the learned average value, usually describe the Gaussian distribution, as,

$$P(\{\oint_{x,y}^{n} Px \,|\Delta\} = \sum \mathrm{C}(\partial_i^{cl}\tau_i^{cl}) \ldots\ldots\ldots (3)$$

In which Δ are the observations, $\oint_{x,y}^{n} Px$ is the pixel feature atx, y, cl denotes the cl-th facial color model, $\partial_i^{cl}\tau_i^{cl}$ are the models parameters learned in training procedure. In the case of skin color features of facial points, the RGB properties are usually converted into YC_rC_b format as it is an existing MATLAB function, thus saves computational aspect and RGB may fail if the lighting condition changes. Because Y is closely related with surrounding illuminations, only C_r and C_b are usually contains chrominance information features for the evaluations. These techniques usually use color components in the color space, such as the HSV or YIQ formats instead of RGB format for input color.

Bayesian description suggests the event of a pixel appearance follows a joint distribution of the neighborhood pixel. In joint distribution with above joint inference, the single isolated face like point will be precluded from the probabilistic reasoning. In total, the likelihood of a region $\aleph_k$ as a collection face-like N points can be tentatively given as the total points

$$f(pl_{(k,i)}) = P(\emptyset_k|(f_i)_{x,y}) = \frac{1}{N}\sum_{x,y\in\emptyset} P\left(\int_{x_0,y_0}^{i} | \int_{x,y}^{i}, \forall\right) \ldots (4)$$

In which $\emptyset$ is the average probability of all the detected points in the region. In the approach, we incorporate a visual salient region [10] by discovering a regional contrast-based saliency extraction algorithm

$$\aleph_r(dt_1, dt_2) = \sum_{i=1}^{n_1}\sum_{j=1}^{n_2} f(pl_{1,i})f(pl_{2,j})D(pl_{1,i}, pl_{2,j}) \ldots (5)$$

In which $f(pl_{k,i})$ is the probability of the ith color $pl_{k,i}$ among all n_k color in kth regiondt_k. The probability of a color is in the probability density function of the face region in equation (2) is used as the weight for the color to emphasize more color differences between dominant colors.

In the second step of the saliency extraction, spatial information is incorporated introducing by including a particular weight term to increase the effects of face-like region and decrease the impact of non-face like regions. Especially, for any region, the spatially weighted region contrast-based saliency is defined as

$$S(dt_k) = \frac{\sum \ \exp(\aleph_\theta(dt_i, dt_k)}{\delta_s^2} w(dt_i)\aleph_r(dt_1, dt_2) \ldots\ldots (6)$$

where, $\aleph_\theta(dt_i, dt_k)$ is the spatial distance between the regions dt_i and dt_k , and δ_s controls the spatial weighting. Larger value of δ_s reduce the effect of spatial weighting so that contrast to further region would contribute more to the saliency of current regions. The spatial distance between two regions is defined as the Euclidean distance between the centroids.

B. Steps of image similarity detection in Facial landmarks

Facial landmarks are discriminative points in a face, such as an eye centroid, left and right eye corners, nose tip, nostril corners, mouth centroid, and mouth corners. A landmark classifier scans the region of interest in a two-dimensional (2D) or three-dimensional (3D) input image and produces the most plausible points based on some measurement criteria [16].

We assume the detected visual saliency VS_1 beween Image img_1 and VS_2 of image img_2 determines the image similarity. Let $VS_1 = \{G_1, G_2, G_3, \ldots\ldots, G_m\}$ and $VS_1 = \{H_1, H_2, H_3, \ldots\ldots, H_n\}$. We compute the shape distamce for the salient regions between two images and accordingly obtain the image similarity values. For every point in VS_1, we search the nearest neighbor H_{nn} in the set of VS_2 of visual saliency regions, which has the shortest shape distance to VS_1 according to the shape distance. Alternatively, for the visual salient region in other set VS_2m we search the nearest neighbor G_{nn} in the set VS_1 which has the shortest distance to VS_1. Intutively, we we have compared any point in $G_i$in VS_1 with G_{nn} to see if it proves the shape of G_i is still we maintained. Otherwise, it happens the notion of drastic shape deformation and henceforth we are unable to find the right corresponding region VS_2 of H_i. The shape maintained regions between the two images defined above, accordingly measure the similarity between two images. We detect shape deformations of visual salient regions between two images through estimating shape distances, and accordingly compute the image similarity values.

C. Adaboost HAAR Classifier

In the machine learning community, boosting decision trees was introduced to combat prediction by linear combination of trees instead of single trees. The AdaBoost classifier is a parallel classifier combined with many linear weak classifiers. Each weak classifier only focuses on the classification of one dimension in the input feature vector. During the entire training process after the goal is given to the classifier, the algorithm is able to self-adaptively increase the number of weak classifiers so as to improve the overall accuracy rate of the classification and focus on the main features. After a weak classifier is added, the algorithm uses the minimum error to calculate the weight value of this weak classifier and re-adjust the weight value of every training sample and then pass the value to the next newly added weak classifier. Based on the recently added weak classifier, the effect of the overall parallel classifier is improved. The final classifier can be expressed as

$$H(x) = sign(\textstyle\sum_{t=1}^{T} \vartheta_t h_t\,(x)\ldots\ \ldots(1)$$

in which, $h_t$is the weak Haar classifier in the cascaded scheme. Haar features can be classified as three types including center-surround features, edge features, and line features. Integral image technique which yields a fast feature computation is applied to extract Haar-like features for the classifier training [2]. A Haar-like feature consists of two or more vertically or horizontally adjacent rectangular regions, and its value is difference between sums of pixels within these rectangular regions [2]. Value of a Haar-like feature (HF) can be represented as weighted sum of intensities of rectangular regions in a feature as in (1).

$$H(F) = (\textstyle\sum_{i=1}^{R} sign(i).\,w_i\,.\,\mu_i\ \ldots\ \ldots\ \ldots(2)$$

where R is the number of rectangular regions constituting a Haar-like feature and sign(i) is the symbol assigned to the ith rectangular area. w_i is the weight which is inversely proportional to the area of the rectangular region and μ_i is the average intensity of the rectangular region.

D. Face Annotation

Face annotation plays a significant role in our system. A face annotation tag incorporates names for identifying and training any person for the face recognition. To this end, we seamlessly connect face annotation with user's contact address. We currently utilized the Google Contacts Data API to maintain the relationship between faces and people in our system. In this paper, confidence in face detection for image annotation is ensured through color based facial classifier with cascaded HAAR classifier devised by Viola and Jones to reduce false positive [1]. This adoption of fusion scheme allows us to combine geometric feature based Adaboost classifier with color based facial region likelihood classifier. We deployed the open source library developed by Intel for the implementation of computer vision. The Open CV Library pertains the field of image processing and includes an implementation of Haar classifier detection and training. AdaBoost classifier in HAAR cascade classifier forwards some of the not identified areas out of interest areas to region based Decision Tree classifier [11] to check if it is a profile face. Fig. 1 shows the fusion scheme adopted by the paper. The paper combines Adaboost classifier with geometric features and objective-based facial color classifier in the visual saliency map, which works at the image level, feature level, and scoring level separately. The advantage of adopting this scheme overcomes the false region generated by the similarity in geometric components and provides a facial likelihood map in the tiles. In this color scheme, we use skin-similar region and shape-based detection calculation to form Benchmark profile-outline faces. In the fusion system, the identified areas in the AdaBoost are re-checked regarding prototyping facial map at image level and feature level.

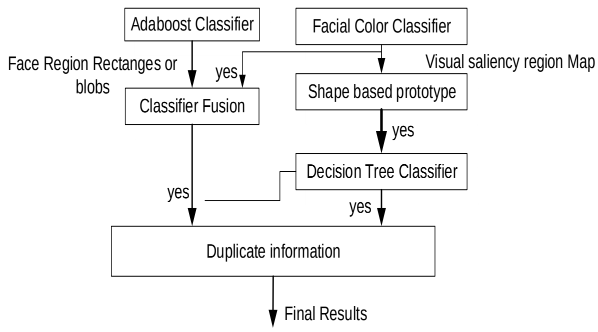


Fig. 1. Fusion Scheme for our classifier

The Adaboost classifier forwards to the particular decision tree classifier to detect the clarity of outline-profile face if the algorithm does not detect the face like the locality of interest. We adopted the framework suggested in [12] which takes a weighted average of color and texture features for DT induction. The advantage of this scheme for the training-based classifier is that it can classify images globally, and we overlay DT to annotate regions of segmented images. One of the significant features of this work is the design of both pre-pruning and post-pruning techniques to train a well-behaved DT.

D. Face region Prediction

After its color features detect a possible face area, we perform region match to average facial features shape of the ellipse to complete the decision diagram. In terms of the comparison between two images, the human visual system in the visual salient regions is very sensitive to the shape variations. Therefore, shape preservations of visible, prominent areas should be taken into account while measuring the similarity between two images. Thus, we may take the geometric structures of objects into account while measuring the image similarity as the same way in geometry based similarity metrics. However, we achieved lower computational cost and better accuracy on visual salient regions when we avoid following the intensity-based similarity measure. We did not establish the correspondences between pairs of pixels since we mainly focus on visual prominent regions. Face contour should be fit into an elliptical shape with an acceptable axis ratio while the face region is detected. We employed Hausdorff distance measure [13] to measure the distance between an ellipse and the shape boundary in the visual salient region. In our work, we use a Gaussian model to evaluate the likelihood P_{shape} of the obtained Hausdorff distance as an acceptable facial component, rather than simply using a distance threshold to reject the detected regions. . If the total likelihood is lower than a user-defined minimum accepted likelihood, the region is rejected. If the region has a face-similar shape, the adopted scheme for decision tree classifier trained by 25 Benchmark profile-outline faces and 25 non-face samples is employed to testify it is a Benchmark profile-outline face.

E. Indexing and Drawing with face position on canvas

In our system, we scaled off the picture by indexing it with the width and the height and then tags it with $\langle user, location\rangle$ tag. In such a tag, user is user name and location is the face position, such as the x and y coordinates from the left-top edges of the photo.

Haar features can be the centroid coordinates of each annotated face that are further calculated to support a fast interactive search based drawing of people positions. Since every photo has a different size, all pictures in the framework are divided into 4×4 blocks (16 rectangular regions). We call this rectangular part as "blob" in our system [6]. Furthermore, a user can manually annotate the blob and extends the number of blobs to cover up the whole face. However, users would be interested in an approximate position of a face in a photo, rather than an exact location of the face. Therefore, we take into account the region containing the centroid to match a positional query quickly and efficiently.

Currently, the user can formulate their plan by drawing blobs onto the canvas while tagging. The user can bring the canvas to specify the regions of face appearances and add a tag for the person in the registered list in contact API. The user can also drag and scale the areas to change positions and the sizes of the face and correspondingly system frame changes accordingly to form search intention automatically.

By default, the query canvas is segmented into 4×4 regions analogous to photos indexed. This segmented region is called "tiles" in our system. This "tile" holds the centroids of the "blobs" (rectangular part of the image) whenever a user draws the position of the face appearance on the canvas.

Although the canvas interface is convenient to represent users' search goal, sometimes user make an error of drawing face positions what they are likely to find. It is worth pointing out that the user can outline a spatial position with which user can adjust the desired accuracy level.

II. EXPERIMENTAL RESULT

Toward the end, in this paper, we present a performance evaluation of the proposed face annotation method. To test the performance of our method with respect to the number of training images, we trained different models from differently sized subsets of the training data. Two set of images we need to train the classifier. One set contains an image or scene that includes the target object. We refer the set of images as benchmark face images. The other part of the database, the negative images may contain scene other than the target image. Face database usual possess the grayscale images. We use 200 set of images to train the Adaboost classifier and selected 25 Benchmark outline view of face images. We also retrieve photos from the popular photo-sharing websites such as Flickr and Picasa from 19 different subjects who are the member of these sites. The objective is to provide color face detection, and correspondingly color image database includes the test bench. In image auto-annotation, the total correct annotation rate is 90% making the fusion scheme useful.

We employed the test images by the color and shape format. We then pass the test image singled out by a measured square window to outline matching algorithm of a picture. We convert the test image to grayscale at the advent of image matching process as described earlier.

TABLE I. FACE DETECTION RESULT OF TEST IMAGES

Training	*numFaces*	*numHit*	*numRepeat*	*numFalse*
TR-1.jpg	22	19	0	0
TR-2.jpg	24	24	0	1
TR-3.jpg	25	23	0	1
TR-4.jpg	23	21	0	1
TR-5.jpg	23	22	0	0
TR-6.jpg	22	22	0	3
TR-7.jpg	24	23	0	1

Moreover, it should be scaled by dividing it to the average brightness of the picture. This scaling would eliminate the effect of the brightness of the test image in the process of image matching. We derive this illumination component as the color outline of the by Facial color modeling in section A. Table 1 shows the result. In here, numFaces= Number of faces, numHit= Number of faces successfully detected, numRepeat=Number of faces repeatedly identified, numFalse= Number of cases in which cases report a non-face. The face detection algorithm shows nearly 93.1% hit rate, 0% number of Repeat and approximately 4.1% false heat rate. In the experimental picture of figure 2, 5 out of 5 images have been detected, and there was no repeat or false detection. Table 2 shows that experimentation used different combinations of the number of training images and the number of initial shapes for each labeled face images. the partial annotation of data. 100 trained examples are fully annotated, and the rest is only partially annotated.

TABLE II. AUGMENTATION OF EXAMPLES WITH NO ANNOTATION AND PARTIAL ANNOTATION OF SETS IN THE RECTANGULAR REGION

Examples	100	200	500	1000
Early Shapes	300	200	80	40
Error	.065	.056	.052	.049
Examples	100	100	100	100
Partially annoted images		1600(25%)	1600(50%)	1800
Error	.075	.069	0.67	0.048

Observation shows that unsuccessful detection of the Benchmark profile-outline faces is segmented, and shape likelihood evaluates segmented blobs. However, show this type of augmentation does not adequately compensate for a lack of annotated training images. Though the rate of improvement gained by increasing the number of training images quickly slows after the first few hundred images. However mixing it with partially labeled data in a third and fourth row shows that we can gain substantial improvement by using partially marked data.

Correspondingly the possible blobs are forwarded to Benchmark profile-outline face classifier trained by a small database set. In a word, this fusion scheme leads reduced false region and hence increases the number of undetected blobs as annotated image. It shows that the spatial position of the centroid that has the highest accuracy of the search empirically, whereas the centroid position increases search relevance. We are currently planning more aesthetic aspect of the image and employing the ranking. Fig. 2 demonstrates the detection of the Benchmark profile-outline faces. The Adaboost classifier failed to detect the outline view faces of a profile in the case of small Benchmark profile-outline database. The possible Benchmark profile-outline face classifier included possible tiles and trained by a small database set. Then the segmented face-like region is evaluated by its shape likelihood. Fig. 2 shows that the Adaboost Haar classifier identified two false region due to their similarity in their geometric components. We get the correction of the wrong detection by inspecting their face likelihood map in these rectangles according to the (4). We eliminated the falsely detected regions from the probable face appearance candidates, precluded by their total color likelihood, as shown in right column image in fig. 3. From the experimental results, it is empirically suggested that the fusion approach can efficiently reduce the false positive detection rate and successfully performed the face detection.

Fig. 2. PFS front ends for auto-annotation. The initial result detected by AdaBoost Algorithm.

The initial experiment shows the fusion scheme can be useful to supply a reliable face classifier for image annotation.

Fig. 3. Adjusted result after classifier fusion.

III. CONCLUSIONS

We have described the development process of a photo frame system with an intuitive, user-friendly interface and in manual and auto-annotation mode. Our system is simple and easy not only to maintain personal memories by also to form a positional query. By drawing the position with tags, we seamlessly connect spatial information with people in photos and enable more flexible photo lookup, according to the user's intention. The developed algorithm needs more features and deployment of the sophisticated algorithm to use it with the more general application. We intend to use more efficient machine learning scheme to improve the HAAR features as well as introduce algorithms for Facial landmark localization.

REFERENCES

[1] Viola, P., and Jones, M., "Rapid object detection using boosted cascade of simple features," *IEEE Conference on Computer Vision and Pattern Recognition*, in *Proc. CVPR'01*, vol. 1, pp. I-511, 2001.

[2] Wilson, P. I., Fernandez, and J. (n. d.), "Facial feature detection using Haar classifiers," *Journal of Computing Sciences in Colleges* 21, no. 4, pp. 127–133, 2006.

[3] R. Feraud, O. I. Bermier, "A fast and accurate face detection based on Neural Network," Pattern Analysis and Machine Intelligence, *IEEE Transactions on* 21, no. 1, pp. 42–53, 2001.

[4] L. cao, J. Luo, and T. S. Huang, "Annotating photo collections by label propagation according to multiple similarity queues," in *Proc. MM'08, ACM International Conference Multimedia*, pp. 121-130, Oct. 2008.

[5] A. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-based image retrieval at the end of the early years," *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 22, vol. 23, no. 1, pp. 1349-1380, Jan. 2001.

[6] Zhang, D., Islam, M.M., & Lu, G. A review on automatic image annotation techniques. Pattern Recognition. 45(1). 346-362, 2012.

[7] H., Xu, J. Wang, X. S. Hua, and S. Li, "Image search by concept map.," in *Proc. SIGIR'10*, pp. 275-285, ACM, 2010.

[8] Kim, H., Saddik, A. El, Edward, K., Lee, K., Lee, Y., & Jo, G. (2011). Photo Search in a Personal Photo Diary by Drawing Face Position with People Tagging, In Proc. 'MM 11, pp. 443–444.

[9] Zhang, C., and Z. Zhang. "A survey of recent advances in face detection." Tech. rep., Microsoft Research, 2010.

[10] Rosin, P. L. (2009). "A simple method for detecting salient regions." Pattern Recognition, 42(11), 2361-2371.

[11] Y.Liu, D. Zhang, G.Lu, "Region-based image retrieval with high-level semantics using decision tree learning," Pattern Recognition 41(8), 2008, pp. 2554-2570.

[12] Sun, L., Ge, H., Yoshida, S., Liang, Y., & Tan, G. "Support vector description of clusters for content-based image annotation,". Pattern Recognition, 47(3), pp.1361–1374, 2014.

[13] Qiang, H.Q., Chun H.Q., and Sheng R. G., "Similarity Measure for Image Retrieval Based on Hausdorff Distance," Applied Mechanics and Materials. Vol. 635. 2014.

[14] Papageorgiou, C.P., Oren, M. and Poggio, T., "A general framework for object detection," In *Computer vision, 1998. sixth international conference on* (pp. 555-562). IEEE.

[15] Hong, S., Khim, S. and Rhee, P.K., 2014, "Efficient facial landmark localization using spatial–contextual AdaBoost algorithm," *Journal of Visual Communication and Image Representation*, *25*(6), pp.1366-1377.

[16] Belongie, S., Malik, J. and Puzicha, J., "Shape matching and object recognition using shape contexts," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *24*(4), pp.509-522.

Automatic Electrical Home Appliance Control and Security for disabled using electroencephalogram based brain-computer interfacing

S. Paul, T. Sultana, M. Tahmid
Electrical & Electronic Engineering, Electrical & Electronic Engineering, Computer Science & Engineering
School of Science, Engineering and Technology, East Delta University
Chittagong, Bangladesh
Email: [jishu.astro, tanni.tanin, badhon434]@gmail.com

***Abstract*—The development of Brain Computer Interface (BCI) system helps to utilize Electroencephalography technology providing an effective way of turning human thoughts into actions as well as communication to other physical devices without any help of the traditional muscular pathways. The BCI system incorporated with EEG technology has recently become a wonderful solution to provide direct communication and interaction pathways for old age persons, sick patients, and especially for the severe handicapped person. In this Paper we propose a novel automatic electrical home appliance control system model using BCI. The proposed system will collect brain signals through EEG equipment and process using a microcontroller. The extracted brain thought signals will further be sent via available wireless communication technology to the input of the receiver microcontroller that will directly activate and control the electrical appliances. The system will be incorporated with a security alert subsystem along with the control purpose where the user can activate an instant alarm by his thoughts in case of danger as well. It is expected that the advantage of portability and cheapness of this proposed system compared to the other BCI systems will make it superior and more user-friendly.**

***Index Terms*—BCI, EEG, Home Automation, Microcontrollers, Wireless, Bispectrum.**

I. Introduction

Motor imagery represents the result of conscious access to the content of the intention of a movement. It can be defined as a dynamic state during which an individual mentally simulates a given action. This type of phenomenal experience implies that the subject feels him/her self performing the action. It is usually performed unconsciously during movement preparation. But a very interesting fact is that, conscious and unconscious motor preparation share common mechanisms and they are functionally equivalent. As a result, a clear image of an intended action can be present even without the limb being involved[1].

A brain computer interface (BCI) is a direct communication pathway between the brain and an external device. It is also called mind-machine interface (MMI), direct neural interface (DNI), or brainmachine interface(BMI). It is a communication system for controlling a device, e.g. computer, wheelchair or a neuro-prosthesis, by human intensions, which does not depend on the brain's normal output pathways of peripheral nerves and muscles but relies on the detectable signals representing responsive or intentional brain activities. It transforms mental intentions into control commands by analyzing the bioelectrical brain activity. BCIs can help patients totally losing volitional motor ability but having intact cognition and improve their living standards[2]. A successful BCI system very much depends on the following criteria: i. Ability of the extracted features to differentiate the task-oriented brain states, ii. Efficiency of the methods for classifying such features in real-time[3]. For analyzing BCIs, the brain activity of a patient has to be recorded.

The traditional Electrode system for acquiring brain signal is International 10-20 system that works with the help of 21 electrodes which are placed on the surface of the human scalp. The usage of common Ag-AgCl small disc metal electrodes followed by proper skin preparation, conductive gel etc. can cause discomfort to the human sample for a long term signal acquisition process. Compared to this metal electrode system, user friendly dry electrodes offer more convenient way to EEG technology[4].

EEG measures voltage fluctuations resulting from ionic current within the neurons of the brain. EEG measures the summation of electrical potentials in the form of electric field generated from millions of neurons having same spatial orientation. The electrical field is mainly developed by currents that flow during synaptic excitation of the dendrites[4]. Advantages of EEG are that it is a very low cost technique, non-invasive and recording procedures are comparatively easier.

The proposed Brain-computer interface (BCI) model will provide an alternative method of expressing human thoughts other than the traditional pathways of peripheral nerves or muscles with the help of the communication based on neural activity generated by the brain.

So, the objective of this research is to prepare a real time basic Brain Computer Interfaced model that will be able to activate Electric loads along with an alarm especially for the disabled people using the classification of EEG signals.

II. General System review

A. General BCI Diagram

The General BCI system consists of Signal Acquisition block, Feature extraction followed by analysis of the signals, classification of the signals and interfacing with the real time machine through computer with the help of machine learning algorithm.

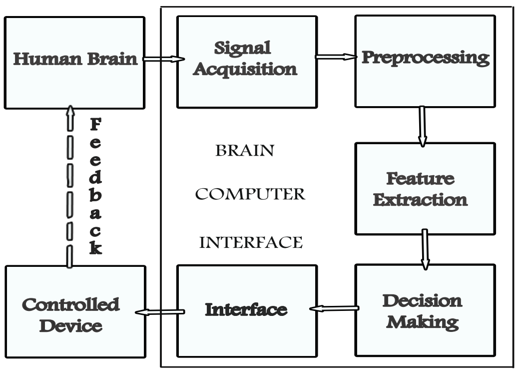


Fig. 1. General BCI system

B. EEG Signal Acquisition system

The signal acquisition part is performed with the help of either wet electrode system or dry electrode system. International 10-20 system, which is a standardized system for electrode placement[5] uses 21 electrodes that are placed upon the human scalp as Fig 2.

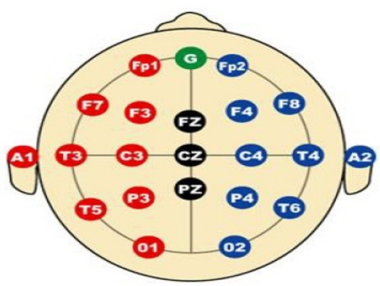


Fig. 2. The 21 Electrode placement of International 10-20 system

The numbers 10 and 20 refer to percentages of relative distances between different electrode locations on the skull. The general electrode- skin interface diagram is as the Fig 3.

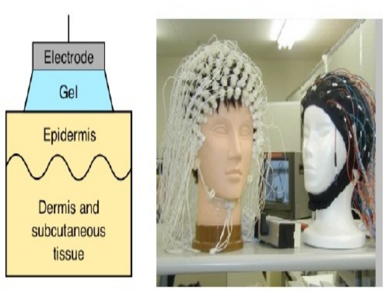


Fig. 3. Skin- Electrode interfacing in the wet electrode system

On the other hand, the newly invented single channel prototype called NeuroSky Mind Wave sensor which is a dry type sensor is able to acquire different brain activities in a non-invasive manner. Compared to the International standard electrodes, brain wave sensor provides less complexity along with increased accuracy. Every NeuroSky product encloses a Think Gear chip which enables the interface between users brain and the load activation unit[6]. This TGAM (Think Gear) module consists of an onboard chip that filters the electrical noise by processing the data sets. Raw brainwaves and the essence (like attention, Excitement and concentration) values are determined[7].
The sensor (Fig 4) performs extraction of raw EEG signal in a non-invasive manner and transmits wirelessly to the processing unit through RF transmitter. Sensor composed of headset, an ear clip and sensor arm[8]. Specifications of such devices are: weight 90g, frequency ranges from 2.42 2.472 GHz and its maximum power is 50mw. Sampling rate of EEG signal is 512Hz.

Fig. 4. The Neurosky mind wave sensor model

C. EEG rhythms and waveform

The recorded EEG signals from the scalp having amplitudes ranging from 20 to 100 microvolts and a frequency content ranging from 0.5 to 30-40 Hz can be conventionally classified into five different frequency bands as the table below:

Table 1: EEG Rhythms at different frequency Range

Signal Type	Frequency Range (Hz)	Observation Peak Time during	Unusual during
Delta	<4 Hz	deep sleep	awake, normal adult
Theta	4-7 Hz	drowsiness and in certain stages of sleep, in infancy and childhood	awake adult
Alpha	8-13 Hz	relaxed and awake with eyes closed	eyes are open
Beta	14-30 Hz	resisting or suppressing movement, or solving a math task	Deep sleep
Gamma	>30 Hz	associated with attention, perception, and cognition	Sleep, subconscious

D. EEG Data Analysis

Signal processing methods can be divided into two general categories: methods developed for the analysis of spontaneous brain activity and brain potentials which are evoked by various sensory and cognitive stimuli. The alteration of the ongoing

EEG due to the stimuli is called event related potential (ERP), in the case of external stimulation called evoked potential (EP). There are mainly three modalities of stimulation: auditory stimuli are single tones of a determined frequency, or clicks with a broadband frequency distribution. Visual stimuli are produced by a single light or by the reversal of a pattern as for example a checkerboard. Somatosensory stimuli are elicited by electrical stimulation of peripheral nerves.
The power spectral analysis provides a quantitative measure of the frequency distribution of the EEG at the expense of other details in the EEG such as the amplitude distribution and information relating to the presence of particular EEG patterns. Hence timefrequency signal-processing algorithms such as discrete wavelet transform (DWT) analysis are necessary to address different behavior of the EEG in order to describe it in the time and frequency domain. It should also be emphasized that the DWT is suitable for analysis of non-stationary signals, and this represents a major advantage over spectral analysis.

III. Methodology

The methodology is divided into two sections. The hardware configuration and the software processing part. First, we explain the proposed hardware model and then show the algorithm used in processing of the brain signal with the simulation results included in the following section.

A. Hardware Model

1) Data Acquisition: Unlike the electrocardiogram (ECG), EEG has a very low amplitude (5-500 uV) and their noisy nature make it hard to detect them. Another issue is the DC offset of the signal due to electrode-tissue interface. This DC offset is usually 20-50 mV and about 500 times bigger than the signal. Thus, a very low noise, high input impedance and high CMRR (Common Mode Rejection Ratio) instrumentation amplifier is required to amplify these signals and reject the DC offset[9]. The following block diagram shows the data acquisition system.

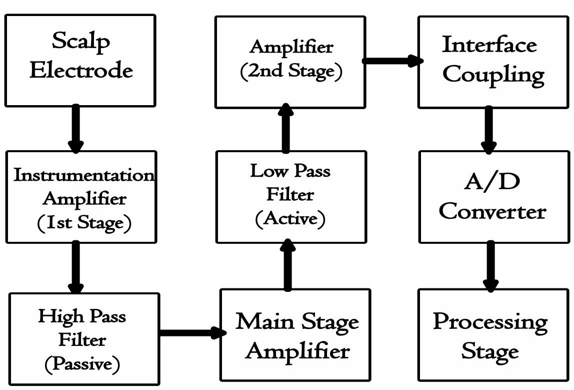


Fig. 5. Block Diagram of the EEG acquisition system

Here, we propose the international 10-20 system of leads as the electrode system.

2) Classification of data in the Microcontroller: The acquired and preprocessed data is considered to be the input to the microcontroller. Here the software part of the system will be in use. A prestored matrix containing the feature vector for the previously stored training data will be used by the machine learning algorithm. The acquired EEG data will be processed according to the algorithm described in the software part to produce the test feature vector. The machine learning algorithm wil be implemented in the microcontroller using the following block diagram for KNN classifier[10].

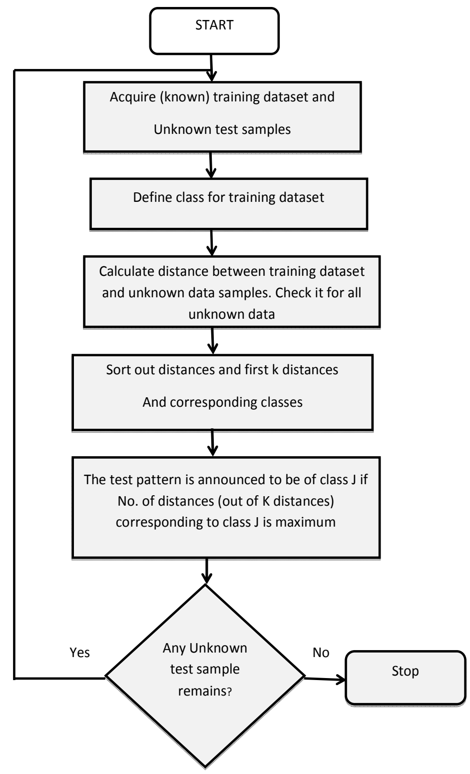


Fig. 6. Block Diagram of the KNN algorithm

3) Wireless Data Transmission: After the processing of the brain signal, the acquired binary result is transmitted through a bluetooth module to the desired receiver device. In our system, we are currently controlling only one load and there are two instructions, namely ON or OFF. But the system can be upgraded easily for multiple loads at the same time.

B. Software Model

We are proposing a classification algorithm which uses the statistical analysis of the bispectrum of the EEG signal. The block diagram in Fig 8 shows the algorithm used in this system. For a test run, we have used an available dataset which features the time delay between the two datasets. It is to be noted that, a real time BCI based system always faces the problem of time difference between training and testing data and most of the existing EEG classification algorithms fail

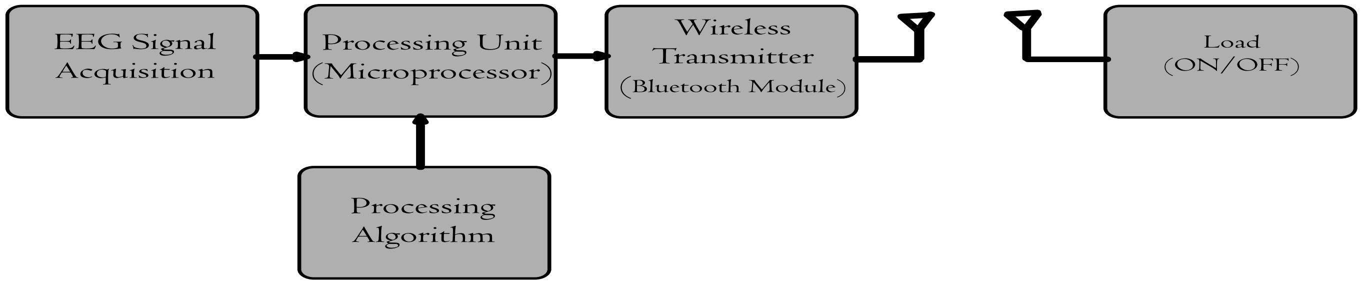


Fig. 7. Overview of the Proposed System

to perform well in this situation. Our proposed algorithm has performed well in this circumstances (checked by simulation).

In our system, we will use two imagery motor tasks to detect the corresponding ON or OFF signals. Corresponding to each instruction a 1 or a 0 is generated.

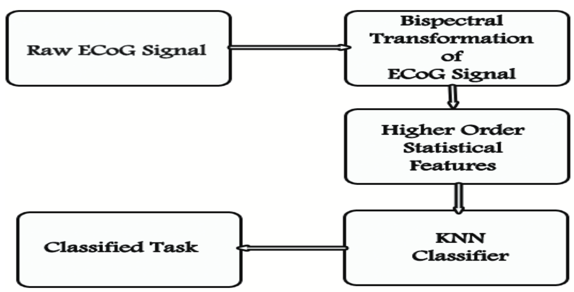


Fig. 8. Proposed Algorithm

1) Details of the Dataset: The dataset used in this research is the dataset I of the BCI competition III, it contains data of imagined motor movement of left small finger or tongue. The train set and the test set were recorded from the same subject in two different days with one week in between. In the BCI experiment, a subject had to perform imagined movements of either the left small finger or the tongue. The time series of the electrical brain activity was picked up during these trials using a 8x8 ECoG platinum electrode grid which was placed on the contralateral (right) motor cortex. The grid was assumed to cover the right motor cortex completely, but due to its size (approx. 8x8cm) it partly covered also surrounding cortex areas. All recordings were performed with a sampling rate of 1000Hz. Further details about the dataset I in the BCI competition III can be found in [11].

After amplification the recorded potentials were stored as microvolt values. Every trial consisted of either an imagined tongue or an imagined finger movement and was recorded for 3 seconds duration. To avoid visually evoked potentials being reflected by the data, the recording intervals started 0.5 seconds after the visual cue had ended. The labeled training data from the first session was stored in a file called Competition_train.mat. It consists of two parts: Part 1: the brain activity during 278 trials. This part is stored in a 3D matrix named X using the following format: [trials electrode channels samples of time series]. Part 2: the labels of the 278 trials. This part is stored as a vector of -1 / 1 values named Y. The unlabeled test data is also stored in a file called Competition_test.mat. It contains 100 trials of brain activity in matrix X (3D format is the same as described above) but it contains no labels Y.

2) Feature Extraction: In this method, we have used bispectrum of the original ECoG signal and extracted some higher order statistical features. The feature vector consists of bispectral higher order statistical features of the whole signal. Bispectrum is a higher order statistical analysis. It is an ideal tool to investigate non-linear interactions. The Fourier transform of the third-order cumulant is called the bispectrum. For a non-Gaussian random process $x(t)$, its third-order cumulant is defined as

$$C_{3x}(m,n) = E[x(k).x(k+m).x(k+n)] \tag{1}$$

Where, E is the expectation of the of the multiplication of the process and its two lagged versions. The bispectrum of this process is defined as the 2D Fourier transform of the cumulant,

$$B_x(\omega_1,\omega_2) = \sum_{m=-\infty}^{\infty}\sum_{n=-\infty}^{\infty} C_{3x}(m,n).e^{[-j2\pi(m\omega_1+n\omega_2)]} \tag{2}$$

On the other hand, if the process is Gaussian, then its third-order cumulant is,

$$C_{3x}(m,n) = 0 \tag{3}$$

That is, any Gaussian noise in the system is nullified by the bispectral analysis. We can define the brain signal as a sum of non-Gaussian random process $x(t)$ and Gaussian noise $w(t)$.

$$z(t) = x(t) + y(t) \tag{4}$$

Then, the third-order cumulant of the signal is,

$$C_{3z}(m,n) = C_{3x}(m,n) \tag{5}$$

In this method, we have extracted 4 features from the calculated bispectrum from each channel of a single trial. All of these are statistical features. The following features were considered,

a) The sum of the logarithmic amplitudes of the bispectrum,

$$F_1 = \sum_{\omega_1,\omega_2\in F} \log(|B_x(\omega_1,\omega_2)|) \tag{6}$$

b) The sum of the logarithmic amplitudes of the diagonal elements of the bispectrum,

$$F_2 = \sum_{\omega \in F} \log(|B_x(\omega, \omega)|) \tag{7}$$

c) The 1st order spectral moment of the amplitudes of the diagonal elements of the bispectrum,

$$F_3 = \sum_{k=1}^{N} k.\log(|B_x(\omega_k, \omega_k)|) \tag{8}$$

d) The 2nd order spectral moment of the amplitudes of the diagonal elements of the bispectrum,

$$F_4 = \sum_{k=1}^{N} (k - H_3)^2.\log(|B_x(\omega_k, \omega_k)|) \tag{9}$$

So, the feature vector is formed as,

$$F = [F_1 \quad F_2 \quad F_3 \quad F_4] \tag{10}$$

3) Feature Quality: We have verified the qualities of the proposed features. The quality is determined by two parameters: 1. Inter-class Separability, 2. Intra-class Compactness. These two parameters for each of the features are shown here. Inter-class Separability: The following figures show the inter-class separabilities for the 4 features extracted from the bispectrum of the signal.

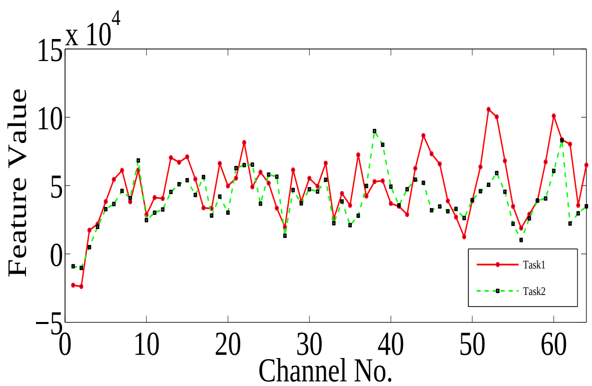


Fig. 9. Inter-class separability of feature F_1

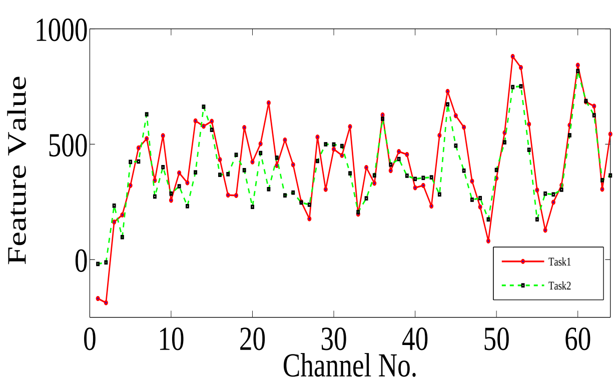


Fig. 10. Inter-class separability of feature F_2

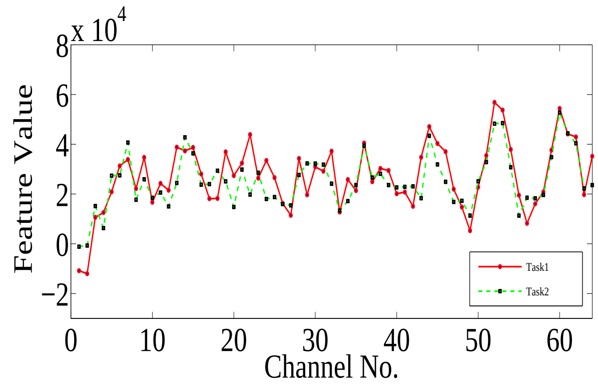


Fig. 11. Inter-class separability of feature F_3

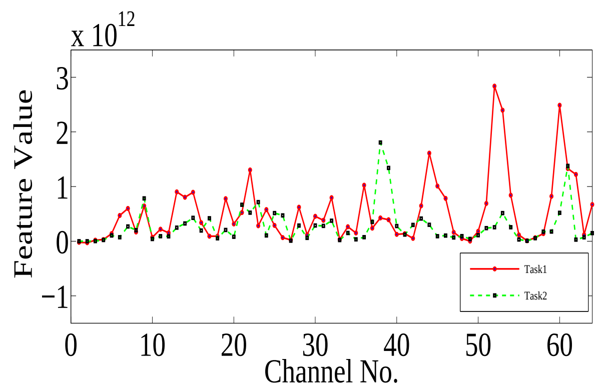


Fig. 12. Inter-class separability of feature F_4

So from the above figures, it can be easily seen that, the feature values of each channel differ by a significant amount for each task. The red line shows feature values for task-1: movement of left small finger and the green line shows feature values for task-2: movement of the tongue.
Intra-class Compactness: The following figures show the compactness of the feature (taking the second feature as example) for each of the two tasks. The black line in the figure represents the mean of the feature values for each channel.

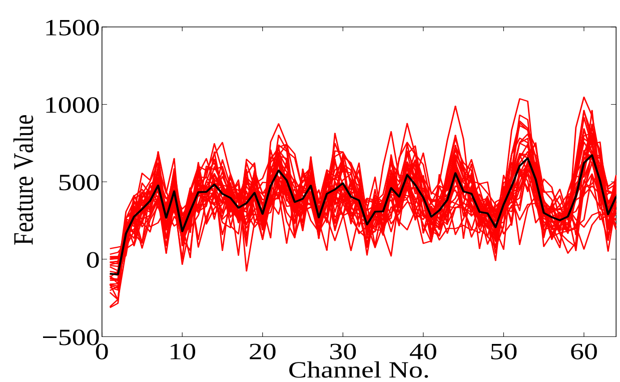


Fig. 13. Intra-class compactness for task-1

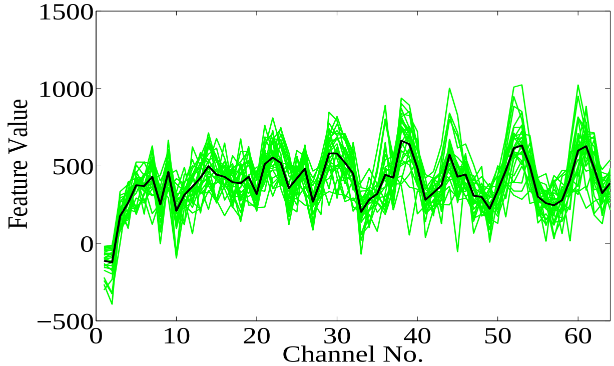


Fig. 14. Intra-class compactness for task-2

From the above 2 figures, the compactness of the features can be seen for individual tasks. With the exception of 2 or 3 channels in some of the trials, all the feature values form a compact band for each individual task.

4) Classifier: In our method, we have used KNN classifier to predict the labels of the test set. K-nearest neighbor is a non-parametric method used for classification. Its input consists of the k closest training examples in the feature space. The output is a class membership. An object is classified by a majority vote of its neighbors, with the object being assigned to the class most common among its k nearest neighbors. It is among the simplest of all machine learning algorithms. In KNN, k is a positive integer. The value of k depends upon the data. Generally, larger values of k reduce the effect of noise on the classification, but make boundaries between classes less distinct. Smaller values of k make boundaries distinct and computations easier; but noise will have a higher influence.
There are three tie-breaking rules: nearest, random and consensus. In our method we have used the first two rules to check the accuracy. The distance metrics are of five types: Euclidean, Cityblock, Cosine, Correlation and Hamming.

IV. Results

We have used accuracy as our performance parameter. Accuracy is defined as the number of labels predicted by the proposed method, matching the true labels provided by the dataset providers.

Though the maximum accuracy falls for some cases, we still get 87% accuracy for the cosine distance metric with nearest rule. Moreover, the optimum value of k is reduced to a very low value of 6. Thus the proposed feature vector can provide a high accuracy with low classification complexity.

V. Conclusion

This research paper showed detailed overview of the general EEG based BCI systems along with a proposed model that will implement the system with the help of the microcontroller. A relatively high performing and real-time efficient classification algorithm is presented here. Moreover our designed and proposed system has the advantage of single processing unit compared to most of the existing systems. Our system is easily upgradable to multiple load capacity without the increase of the number of processing units. Most of the existing systems include the processing unit with the load, hence increasing number of loads becomes tedious and costly. Also, we have designed the processing algorithm keeping in mind the time-varying and random nature of EEG signals. Though some further future expansions are needed, the proposed system surely has the qualities to be an effective and efficient automation system to be useful for the handicapped.

References

[1] Martin Lotze and Ulrike Halsband, *Motor imagery*, Journal of Physiology - Paris 99 (2006) 386-395.

[2] Wolpow JR, Birbaumer N, McFarland DJ, et al. *Brain-computer interface for communication and control*, Clinical Neurophysiology, 2002, 113: 767-791.

[3] S.M. Zhou, John Q. Gan, F. Sepulveda, *Classifying mental tasks based on features of higher-order statistics from EEG signals in braincomputer interface*, Information Sciences 178 (2008) 1629-1640.

[4] Lukas Maly, *WHEELCHAIR CONTROL USING EEG SIGNAL CLASSIFICATION*, MS Thesis Paper, BRNO UNIVERSITY OF TECHNOLOGY.

[5] Jasper, H. H., *The ten twenty electrode system of the international federation*, Electroencephalography and clinical neurophysiology, 10 (1958), 371375.

[6] Athanasios Vourvopoulos, Fotis Liarokapis, *Evaluation of commercial brain-computer interfaces in real and virtual world environment: A pilot study*, Computer and Electrical Engg. 40(2014) 714-729.

[7] L.R. Stephygraph, N.A. Kumar, V. Venkatraman *Wireless Mobile Robot Control through Human Machine Interface using Brain Signals*, International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM), India. 6 - 8 May 2015. pp.596-603.

[8] NeuroSky related informations available at *www.neurosky.com*

[9] Lin ZHU et al. *Design of Portable Multi-Channel EEG Signal Acquisition System*, Biomedical Engineering and Informatics, 2009. BMEI '09. 2nd International Conference on.

[10] Neerja S. D., Rupesh S. M., *Methods towards invasive human brain computer interfaces*, International Journal of Scientific and Research Publications, Volume 3, Issue 6, June 2013.

[11] Lal TN, Hinterberger T, Widman G *Robotic Automation through Speech Recognition*, Advances in Neural Information Processing Systems, Cambridge: MIT Press, 2005, 17: 737-744.

Modeling a Zigbee and PLC based smart energy monitoring and management system to reduce rolling blackout in Bangladesh

Md. Rakin Sarder[1], Feroz Ahmed[2], Tahsin Ferdous Ara Nayna[3]
Department of Electrical and Electronic Engineering,
Independent University, Bangladesh
Email: tunchunairarko@gmail.com[1], fahmed@iub.edu.bd[2], nayna_08@outlook.com[3]

***Abstract*—The demand and consumption of energy is increasing day by day in the urban areas of Bangladesh due to the radical growth and developments in the urban system. The major concern is that even though generation of electrical energy has increased significantly, the rolling blackout phenomenon is still not decreasing in Bangladesh. This paper proposes a method of reducing the rolling blackout phenomenon in the urban regions of Bangladesh. Our proposed method involves implementing a smart energy monitoring and management system which would work as a subsystem of the existing electricity distribution network using zigbee and PLC based network, where the central control unit is situated in different substations across the country. The electric loads, equipment and appliances have been classified into different layers basing upon consumption and necessity. A zigbee based sensor network will perform the metering and monitoring of the power consumption of the building and other structure. The substations central control unit acquire the required data via power line communication (PLC), and control loads based on layer classified. The proposed model ensures that the area never goes to complete blackout, and the basic electric demand of the consumers is fulfilled.**

I. Introduction

The emergence of innovative technologies have enhanced the living standards and quality of life in the modern age. Electricity and electrical devices are a significant part in the modern lifestyle. But such upsurges in the ingesting of electronics and electrical appliances have unfavorably resulted in an unprecedented escalation in energy consumption[1].

In Bangladesh the scenario is quite similar. The generation of electricity and its supply was very low in the past. For instance, in 2003 nearly a total of 2,428 MW power was supplied across the nation, though the demand for electricity was much less that time, about 3,944 MW[2]. The demand was low at that time because the access to electricity was ony 32% of the total population[2]. The generation of electricity and its supply has increased significantly in the recent years, compared to the past. The average generation of electricity in Bangladesh is now about 8177 MW[3]. But the demand for electricity has risen to 11,405 MW, even though 75% of the total population has now the access to electricity[3]. This shows that even though the generation has increased, the demand-supply gap has yet to be neutralized.

Due to the demand-supply gap, load shedding percentage has still not been decreased significantly. Bangladesh has to go through rolling blackout load shedding scheme. This demand response is necessary because the demand for electricity sometimes heavily surpasses the power supply capability of the nations current network[4]. Owed to the less power supply capability, the frequency of the AC voltage in the transmission and distribution line decreases. Such decrease results in reactive drop in the transmission line. As a result, the substation has to create load shedding in its controlling area, in order to balance the supply-consumption level. In Bangladesh, the power line frequency is 50 Hz. As per the national standard for load shedding, if the frequency is decreased by a minimum 0.5 Hz in the area, the controller of the substation should initiate load shedding.

Also, one of the other reasons of rolling blackout is low generation capability. During the time of low generation capability, the power stations passes message regarding the generation to the substations. Low power generation can be for various reasons. For example temporary closure of power plants, failure in transmission lines, natural calamities etc. Some of these situations requires mandatory load shedding across the electrical network.

Load shedding affects Bangladesh's economic and residential sector of the urban areas heavily. A significant portion of the urban areas runs with almost basic electric consumption. Due to load shedding, people residing or working in these areas are the worst sufferers, they are being deprived of their basic electricity need. Also due to inefficient management the load in the generation and distribution of electricity leads to widespread load shedding, resulting in severe disruption in various economic activities, industrial production, education, medical services and so on. It is estimated that power outages result in a loss of industrial output worth $1 billion a year[5]. Continuation of such huge loss is reducing the GDP growth by about half a percentage point in Bangladesh[5].

While it is quite difficult to change the current infrastructure or to make drastic increase in power generation, it is possible to design integrated advanced power monitoring and control mechanisms with the capability to efficiently monitor and control power consumption, so that the power management unit will have its own intelligence in determining proper distribution, optimization and load balancing, instead of total blackout. This can be an alternative solution by which the country can reduce the rolling blackout, with the existing power system and its infrastructure. Our proposed the model of reducing load shedding phenomenon with smart power monitoring, load control and management and thus maintain stability in demand and supply of electricity using SEMMS.

II. Related Works

Systematic monitoring and management of energy ensures adequate measures to save energy and to reduce rolling blackout. In broader sense implementation of such systems is imminent since the infrastructure of power sector in Bangladesh is quite inefficient. Optimization and management of energy can be started from the residences using zigbee based HEMS [6]which proposes an implementation of an intelligent remote energy management system, based upon solar energy. Baig et. al [7]proposed a Zigbee based HEMS for Monitoring and Scheduling of Home Appliances. The suggested system primarily proposed a software based Energy Management Center (EMC) using LabVIEW. The design of the system is cost effective and has a single consumer centric approach. Suryadevara et al [8] proposed a system which has cost-effectiveness, exibility, and a management system which is real time, based upon Zigbee WSN. Another feature of the proposed system is that it is integrable with the existing home power monitoring system.

An essential part of smart energy monitoring and smart grid context is integration of smart meters in both consumer ends and distributor end [9][10]. The study also showed different categories of smart meters in different sections of the network with automatic power reading and prepaid energy system for smart meter.

Ciabattoni et. al [11] proposed a new system for managing loads called "shiftable loads", which is a load management technique aiming to maximize savings by transferring household electricity demand from PV off-peak hours production of the day to production hours which are peak. Milioudis and Galli et. al [12][13] explained how power line communication (PLC) devices are suitable for efficient integration into the distribution lines for communication for various robust applications.

III. SEMMS-Smart Energy Monitoring and Management System

Smart Energy Monitoring and Management System (SEMMS) is a proposed zigbee and PLC based control subsystem which has the ability to monitor all the power consumption activities in a building or any structure using zigbee network, and transmit the information to the control substation via the power lines where SEMMS's central control unit will decide which over consuming loads should only be shed off over a certain time period, rather than total blackout, including basic consuming loads. The simplified system model is given below:

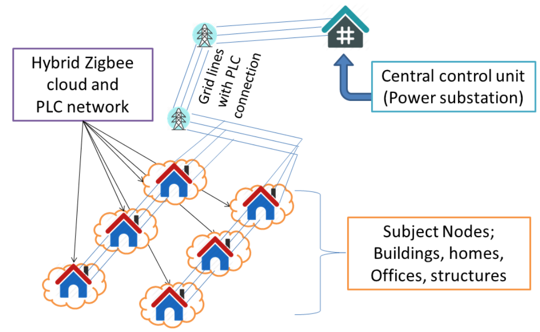


Fig. 1. External system architecture of SEMMS

In SEMMS the central control unit will be the substation. The subject nodes are the buildings, homes, offices and other structures under the distribution area of the substation. The subject nodes meter out the power consumption of pre-defined layers of electrical elements. This metered data is then sent to the central control unit for analysis. When SEMMS gets information that there is a possibility of frequency reduction in the electric line, the system automatically sends command to the subject nodes to switch off the layer classified equipments. When the frequency comes back to normal state of 50 Hz, the central control unit sends command to turn on the equipments back again.

A. Electrical Layer Classification

In order for SEMMS to operate correctly, the most important methodology it needs to follow is Electrical Layer Classification. Electrical layer classification means classifying electricity consuming components and equipments basing upon some certain parameters. Different parameters classifies electrical components in different layers, such as: total consumption, necessity, consumption pattern etc.

In the proposed model, we have modeled four consumption layers based upon total consumption and necessity. These layer classifications are significant because instead of total blackout, SEMMS monitors and manages the layers separately. By managing the individual layers, the system determines which layer needs to be switched off when a possibility of frequency distortion is found. The four layers are:

- Layer 1: Basic consumption layer; Electrical components and equipments which serves the basic needs of consumer
- Layer 2: Living and secondary components; Electrical components and equipments which are required for residential living and working.
- Layer 3: Comfort and casual; Electrical components and equipments which are used for performing various specific tasks that makes living and working comfortable. These components also tends to consume more energy than the previous layer, excluding some exceptions.
- Layer 4: Luxurious and High consuming components; Electrical components which consumes extreme high

amount of energy, about 70-80% of the total consumption. Some of the components falling under this layer falls under the category of luxury.

A table consisting some examples of components of each of the layers is given below:

TABLE: Electrical component examples of layer classified for SEMMS[14]

Layer number	Electricity consuming components
Layer 1	Light sources like bulb, tube light, corridor light; basic cooling component like ceiling fan, table fan
Layer 2	Television, refrigerator, water pump, exhaust fan, radio, computer, device charger, inverter, toaster, blender, printer, sound system etc
Layer 3	Electric iron, washing machine, oven, chapati maker, dessert cooler, water cooler etc
Layer 4	HVAC (heating, ventilation, and air conditioning) components like Air conditioner, Split Air conditioner, Geyser, Immersion heater etc

B. Central control unit

In SEMMS, the central control unit will be set up in the substation. The central control unit consists of server station composed of a clustered computer with a PLC gateway and an electric frequency sensor. The main sections of the unit are discussed below:

1) Clustered computer: The clustered computer is modeled as an ARM architecture processor with 3 Gigahertz processor and 6 megabyte cache, 32 gigabyte RAM, Linux OS, over clocking ability, built-in expandable web server, extension capabilities and appropriate for parallel processing. An ethernet router connects the computers inside the cluster.

2) PLC gateway: For the system modeling, the proposed gateway is a Connected Grid Router suitable for smart grid application. The CGR supports various wired and wireless interfaces, in a modular platform. Supporting a 902- 928 MHz IPv6 RF mesh, CGR can conglomerate up to 5000 nodes, like smart meters. The router can connect to sensors, capacitor bank controllers, recloser controllers, and remote terminal units via integrated Ethernet and serial interfaces. Also for easy integration to legacy (non-IP) devices onto an IP

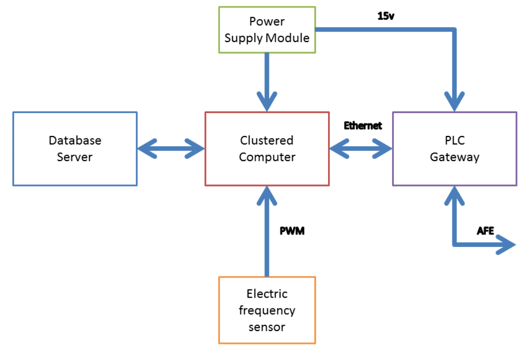


Fig. 2. Block diagram of CCU

network, it follows SCADA protocol (serial-to-IP) translation features. Supporting a variety of communications interfaces, for instance ethernet, serial, cellular, WiMAX, radio-frequency (RF) mesh, and power line communications (PLC), its performance is robust in industrial and grid application[15].

3) Electric frequency sensor: The electric frequency sensor is a sensor which senses both the transmission and distribution line frequency. As stated before, another primary requirement of SEMMS is to sense line frequency. The basic structure of the sensor is made up of a microcontroller with a step down transformer and a potentiometer. The sensor senses the frequency as each iterations of the microcontroller processing and sends the frequency data to the clustered computer.

C. Subject Nodes

In SEMMS the subject nodes which are the buildings, homes, offices and other structures which falls under the managing area of the substation. The main objective of the subject node is smart monitoring/metering electric power consumption of the layer classified loads. An example architecture of a building as a subject node is given in the figure [3]. There are several sections of the subject node. The metering of the power consumption will be done in each of the apartments of the building. Each of the apartment has multiple sensor nodes in it with a small control unit. The control units of each apartments will be connected to the building's control unit.

1) Apartment monitoring and controlling architecture: The monitoring and controlling system of the apartments will be based upon a zigbee mesh network which interconnects the sensor nodes with the control unit of the apartment. The zigbee network has been chosen instead of wired or other wireless networks because:

- Wired network's connections' can be disturbed or disconnected physically with ease.
- Open-source specification
- Unlike star, ring, bus or tree topologies, ZigBee supports mesh network and allow for alternate pathways between

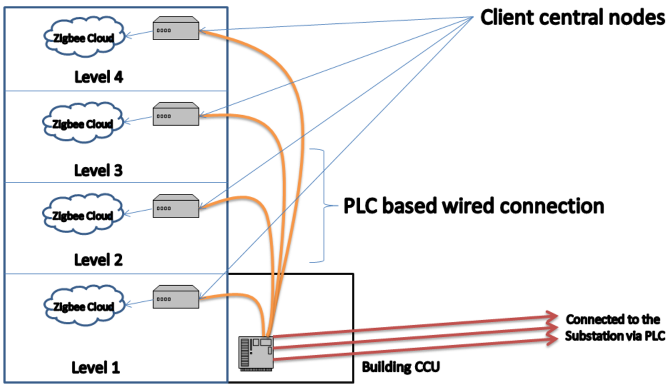


Fig. 3. Architecture of a model subject node

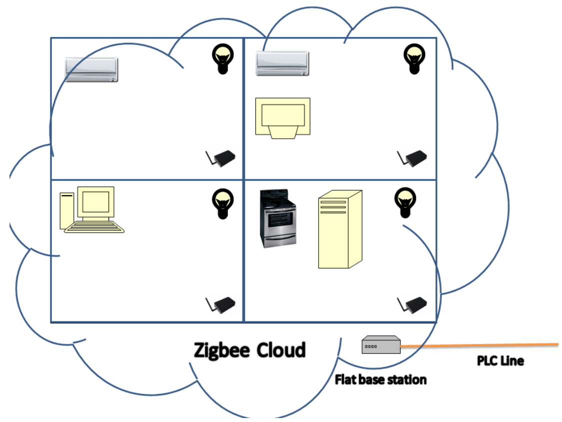


Fig. 4. Architecture model inside an apartment

target devices and controllers. As a result the system becomes scalable, and by keeping inside the range of other devices, a new device can be placed anywhere[16].

- In order to create a mesh network topology using wired connection, separate implementation is necessary, whereas zigbee protocols and wireless are specifically suitable for mesh network setup and is quite easy to set up.
- Adequate data rate with a near real-time response compared to wifi and bluetooth transmission.
- Zigbee is designed particularly for embedded systems with low latency modes, short message modes, broadcast messages etc which is suitable for IoT purposes with low power consumption[17].

Due to above mentioned reasons, zigbee based network has been chosen to integrate the sensor nodes with the apartment control unit.However, since zigbee has short range, low complexity and low data speed, connection between the apartment's control unit and the building's control unit is a PLC based connection. The architecture of the proposed system inside the apartment is given in fig [4].

2) Sensor node: The sensor node will comprise a DC power module, a MCU, an AC power control module and a zigbee module. An MCU communicates with the power measurement module by an analog front end (AFE) and with the ZigBee module through universal asynchronous receiver/transmitter (UART) interfaces. The ZigBee module communicates with the control module by pulse width modulation (PWM) technique. The major functions of the sensor node are as follows:

- Specify electric consumption layer for the node
- Measurement of power parameters, such as the voltage, current, and power of the outlet;
- Control of the power output of the outlet via relay switch;
- Security protection from overload;
- Transmission of the information of each node to the apartment control unit through ZigBee.

The block diagram of the sensor node is given in fig [5].

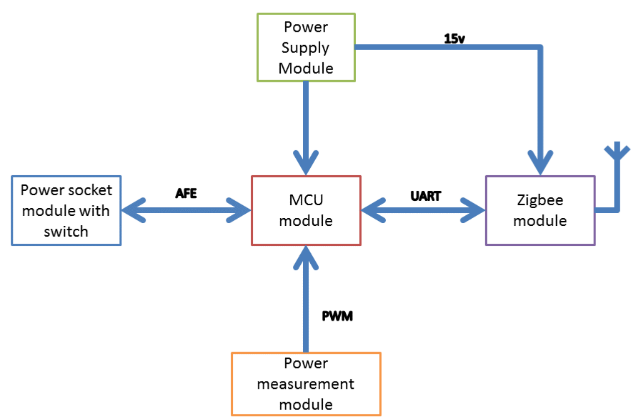


Fig. 5. Block diagram of sensor node

The working flowchart of the sensor node is given in fig [6]. The flowchart contains a simplistic outline of work flow of the sensor node. It also contains how the communication network is established with the Zigbee cloud.

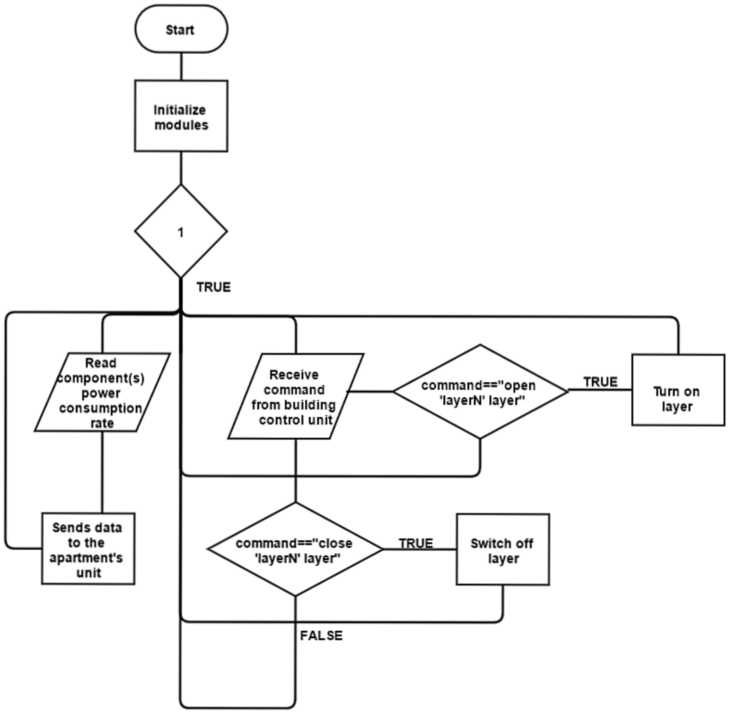


Fig. 6. Sensor node working flowchart

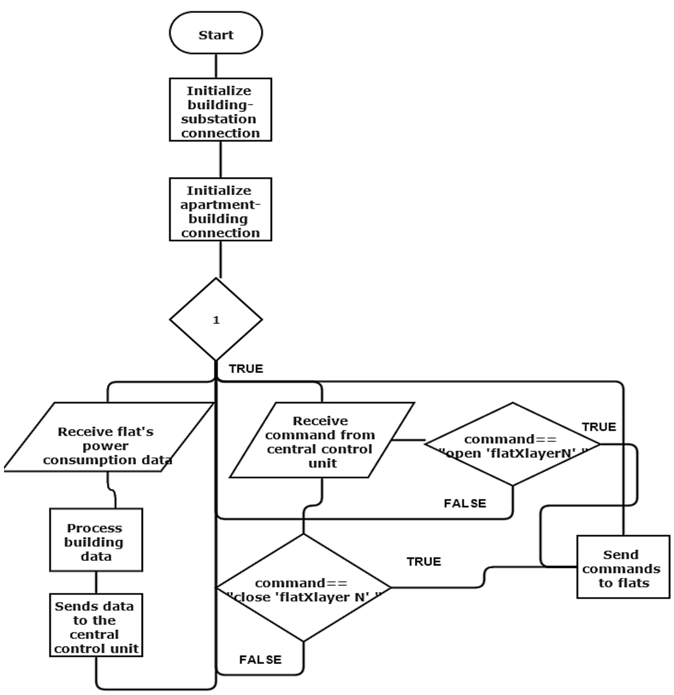


Fig. 7. Building control unit working flowchart

3) Building control unit: The building control unit consists of a DC power module, an embedded home server and a PLC base station. The PLC base station acquires data of each of the apartments' control unit. After that it sends the monitored data to the substation's cnetral control unit. The base station receives analysed feedback from it and perform accordingly. The working flowchart of the building control unit is given in fig [7].

IV. Conclusion

In this paper we have proposed an efficient smart model to solve Bangladesh's national load shedding problem using smart energy monitoring and management. We have proposed a complete subsystem over the existing electrical distribution system in the urban areas of Bangladesh. The study specifically modeled an intelligent electric network that is capable of smartly controlling loads in order to reduce total blackout. A detailed model of smart building including smart apartments have been presented. The apartments are based upon a zigbee based operational and monitoring network, while the building control unit is based upon a PLC network. Also, an efficient method of control over the loads connected to the substation's distribution has been presented. The paper also presents a method of classification of electrical components based upon various parameters.The vast significance of the electric layer control in energy monitoring and management has also been mentioned. By considering both consumption and generation, the proposed architecture is expected to enhance energy management and to mitigate load shedding in urban areas of Bangladesh.

Acknowledgment

The authors would like to thank the Department of Electrical and Electronic Engineering of Independent University, Bangladesh for initiating the long term research.

References

[1] Changhai Peng and Kun Qian, "Development and Application of a ZigBee-Based Building Energy Monitoring and Control System", [The Scientific World Journal, vol. 2014, Article ID 528410], pp. 1-13, 2014. doi:10.1155/2014/528410
[2] Cell, Power. "Bangladesh Power Sector Data Book." [Power Division, Power Cell. Dhaka: Ministry of Power, Energy & Mineral Resources, Government of Bangladesh (2006).] : pp. 4-6
[3] Cell, Power. "Power Sector at a Glance." [Ministry of Power, Energy & Mineral Resources, Government of Bangladesh (2016), http://www.powercell.gov.bd/index.php?page-id=267]
[4] Wikipedia contributors. "Rolling blackout." [Wikipedia, The Free Encyclopedia.] Wikipedia, The Free Encyclopedia, 12 Feb. 2016. Web. 28 Mar. 2016.
[5] VYAPAARASIA. "Power sector in Bangladesh." [Asia trade hub] (2011), http://www.asiatradehub.com/bangladesh/power.asp
[6] Shewale, Archana N., and Jyoti P. Bari. "Renewable Energy Based Home Automation System Using ZigBee." [Renewable Energy 5.3] (2015).
[7] Baig, F., et al. "Smart home energy management system for monitoring and scheduling of home appliances using zigbee." [Journal of Basic and Applied Scientific Research 3.5] (2013): pp. 880-891.
[8] Suryadevara, Nagender Kumar, et al. "WSN-based smart sensors and actuator for power management in intelligent buildings." [Mechatronics, IEEE/ASME Transactions on 20.2] (2015): pp. 564-571.
[9] Yadav, B. Radhika, and J. V. Sharma. "Smart Energy Meter Based, Prepaid Electricity Distribution System." (2015).
[10] Pereira, R., et al. "Consumer energy management system with integration of smart meters." [Energy Reports 1] (2015): pp. 22-29.
[11] Ciabattoni, L., et al. "Residential energy monitoring and management based on fuzzy logic." [Proceedings of the IEEE International Conference on Consumer Electronics, Berlin, Germany.] 2015.
[12] Milioudis, Apostolos N., Georgios T. Andreou, and Dimitris P. Labridis. "Detection and location of high impedance faults in multiconductor overhead distribution lines using power line communication devices." [Smart Grid, IEEE Transactions on 6.2] (2015): pp. 894-902.
[13] Galli, Stefano, Anna Scaglione, and Zhifang Wang. "For the grid and through the grid: The role of power line communications in the smart grid." [Proceedings of the IEEE 99.6] (2011): pp. 998-1027.

[14] "DESCO Appliance Calculator",[https://www.desco.org.bd/index.php?page=bill-calculator]
[15] Cisco Connected Grid 1000 series router datasheet. [Document ID: 1458197357166241 March 16, 2016. http://www.cisco.com/c/en/us/products/collateral/routers/1000-series-connected-grid-routers/datasheet-c78-696278.html]
[16] European Editors. "Understanding the Benefits of Using ZigBee for Wireless Lighting Control", [Digikey Electronics. Contributed By Publitek Marketing Communications, 2013-06-11, sec-3, para-4]
[17] Han, Jinsoo, Chang-Sic Choi, and Ilwoo Lee. "More efficient home energy management system based on ZigBee communication and infrared remote controls." [Consumer Electronics, IEEE Transactions], on 57.1 (2011): pp. 85-89.
[18] Lien, Chia-Hung, et al. "Power monitoring and control for electric home appliances based on power line communication." [Instrumentation and Measurement Technology Conference Proceedings, 2008. IMTC 2008. IEEE.] IEEE, 2008.
[19] Bangladesh power development board report "Present Installed Generation Capacity (MW) as on March, 2016", [http://www.bpdb.gov.bd/bpdb/index.php?option=com-content&view=article&id=5&Itemid=6]
[20] "An Overview of Power Sector of Bangladesh", [Bangladesh power development board, November, 2011] : pp. 9-15

An Android Based Femina Security Alert System

Sayeda Fauzia[1], Shajeda Khanam[2], Sanjida Sharmin[3], Md. Khaliluzzaman[4*] and Rezaul Karim[5]

[1,2,3,4]Dept. of Computer Science and Engineering, International Islamic University Chittagong, Chittagong-4203, Bangladesh

[5]Dept. of Computer Science and Engineering, University of Chittagong, Chittagong-4331, Bangladesh

fauziarafa@gmail.com, jahinshamu0308@gmail.com, ssharmin114@gmail.com, *khalil_021@yahoo.co.in, pinnacle_of_success@yahoo.com

Abstract — **In continuously upgrading world, women belief in their self-worth. They have the same participation like men in almost every sector of life. But lives of women have become so vulnerable these days. Security of their lives is one of the burning questions. Everywhere, they have to face sexual harassment, rape, murder, mug etc. Considering all these harassment towards female in present days, an idea of mobile application came in consideration. In this regard, a Bluetooth security device has been aligned with an android mobile application. Bluetooth security device will be placed in women finger ring. When the button of the ring is pressed, the Bluetooth security device pings the android Smartphone's application to send personalized text messages to the pre selected contacts with the users present GPS location reported from the android smart phone. It also sends a voice call to the one pre defined contact number automatically.**

Keywords—Security; Bluetooth; location; button; application

I. INTRODUCTION

In modern era believes in women empowerment. Women are contributing in every sector of the country. But today female's security at night and even at day time is big concern. Most of the time they are the victims, the criminal stays free. Rape, acid violence, dowry related violence, sexual harassment, mug etc. are very common indecent in our country, which is a very shameful fact. Every day these situations are found in the daily newspaper whose are full of violence against women.

The result of VAW Survey 2011 identified that as many as 87% of currently married women have ever experienced any type of violence by current husband and 77% reported any type of violence faced during the past 12 months from the survey time. The higher percentage of any type of violence is predominantly contributed by psychological violence. Almost 90% of those who have ever violated by current husband has the past 12-month experience of violence which implies the persistence nature of violence by the spouse.

Besides, a survey is conducted to measure the Gender-based violence in [1]. According to this survey, it is seen that in Bangladesh 66% of married women are experienced violence by current as well as previous husbands, while 98% have ever been violated by either current or previous husbands. So, overall survey depicts that the women security has become very essential.

In this section provide a descriptive summary of some technique that have been implemented and tested for women security. The Dhrubajyoti Gogoi et al. [2], proposes an Android Based Emergency Alert Button system which describes about an SOS application developed in android platform. The uniqueness of this application apart from other SOS application available is that the user need not spent time navigating inside the phone menu; unlock the screen, to trigger the service [2].

Jorge Zaldivar et al. proposes a method providing Accident Detection in Vehicular Networks Through OBD-II Devices and Android-based Smartphone [3]. Alexandre Bartel et al introduced an automatically Securing Permission- Based Software by Reducing the Attack Surface: An Application to Android was proposed [4].

Sharma et al. describes an android application which interprets the message a mobile device receives on possible intrusion and subsequently a reply (Short Message Service) SMS which triggers an alarm/buzzer in the remote house making others aware of the possible intrusion [5].

After analyzing all the ideas, an idea is emerged which is very helpful for women for helping themselves. This work consists of an android mobile application and a Bluetooth security device. The individuality of this application apart from other application available is that the user does not need to spent time to unlock the screen, open the application to search inside the contact list to make voice call and write message. At the emergency situations, user can directly press the button of the ring and thereby, a signal from the device will be received by the android mobile. After receiving signal an automatic message of their location information will be sent to selected numbers through the application and also a voice call will be sent to one of the predefined numbers. Many applications and device available which sends a custom message to the predefined number but not the location of the user .This feature of the application not only helps to find the exact location of the person in problem but also will help the nearest police to trace the location of incident immediately.

The paper is organized as follows. In Section II proposed work is described. In the next section performance result is explained. The paper is concluded in Section IV.

II. SYSTEM IMPLEMENTATION

In this section, described the system work flow in details. The flow diagram shows the main target of this work. The working flow is divided into five major parts that are:

1) Connecting the security device with the mobile android application, 2) Security device signal process by the mobile application, 3) retrieve the GPS location, 4) Retrieve the emergency contacts, and 5) finally, message and voice call is sent to the selected contact numbers with the victim's present location. The workflow of the proposed work is shown in Fig. 1.

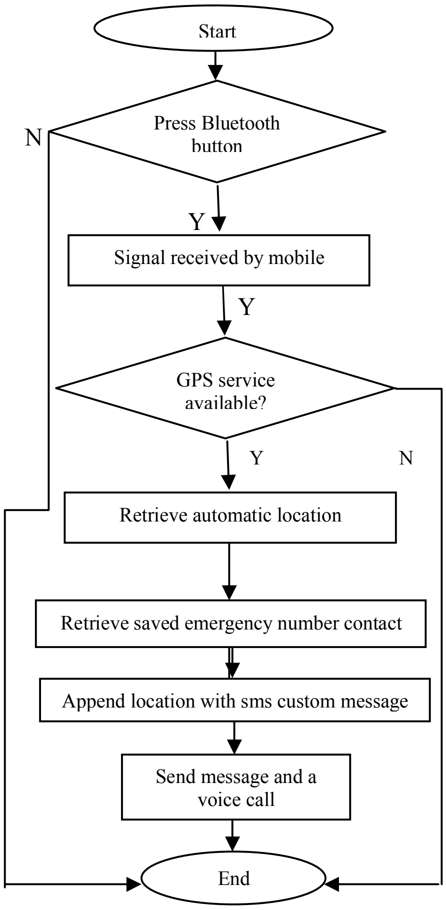


Fig. 1. Workflow of the proposed system.

A. *Connecting security device with the mobile application*

In the initial stage, the security device has to be turned on. Then android mobile need to be discovered the device. And complete the pair to the android cell phone. It requires security code to be paired to make the connection secured. Fig.2 shows the Bluetooth device with security device which sends the signal. The Bluetooth device is in the finger ring. It will be activated when the finger ring is pressed.

Fig. 2. Prototype of the security device.

B. *Security device signal process by the mobile application*

This section describes the procedure of the mobile application that process the signals come from the security device. The security device has a button for sending signals to the mobile application. At the emergency situation, the user will press the security button and it will send a signal to the mobile application. The mobile will receive the signal. After that, the mobile application performs the operation to send the user information to the predefined contact lists. For that, the mobile application has to be retrieved the GPS location of the user android smart phone.

C. *Retrieve the GPS location*

In this section describs the process of retrieving the GPS lication from the user's current locaiton through the android mobiel applcation. For this step, user needs to keep smart phone's GPS on. As soon as, the application will get the signal from the security device which is in the user's finger ring, it will start its main work on it. The application will retrieve the GPS location of the user's present location. Fig. 3 shows result of retriving user's GPS location.

Fig. 3. Retrieve the GPS location of the user's current location.

D. *Retrieve the emergency contacts*

This section described the process of retrieving the specific number of emergency contacts from pre define contact list. After getting the GPS of user's current location, the mobile application will retrieve previously defined emergency contact numbers. User can add as much number as the user want and modify it later on. Fig.4 shows the customized contact list. When user press the security device, the current location of the user will be send to those contacts that are in the top of the contact list. The number of the contact to which the message will be send is defined by the user. However, voice call will only send to the top of the contact number. That means, voice call is passed only one contact number and message will be send a list of the contact number that is defined by the user.

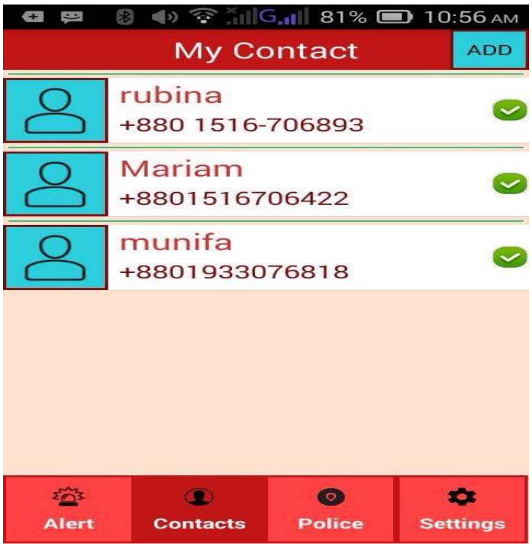


Fig. 4. The customized contact list.

E. Passing message and voice call to the pre difine contacts

This section defines the procedure of sending the message to the pre selected contact list. For that, current location of the GPS is retrieved and the contacts are selected in the previous section. Now, mobile application appends the retrieval user current location GPS link to the predefined alert message. Then the message will be sent to the emergency contacts that are selected previously. User can modify the message also as per user's personal situations.

To access the message with user GPS location, the receiver also needs to keep his mobiles GPS on to get the location and direction towards it. Fig. 5 shows the message received by the receiver of specific contact list. After receiving the user message, the receiver just accesses the link to see the user's current location through the GPS system. From this GPS system the receiver can take necessary action to save the victim.

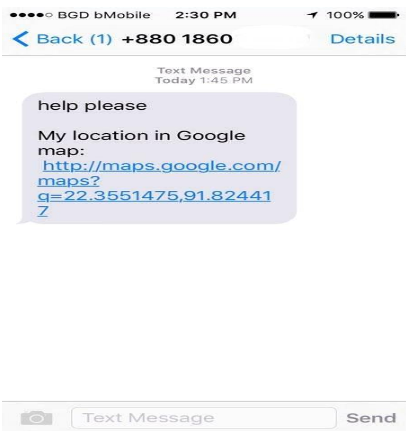


Fig. 5. Sending message to the pre define contact numbers.

Fig. 6 shows the direction towards the victim's current location. According to the figure, the receiver retrieves the current location of the victim after opening the getting message which shows direction of victims via Google map.

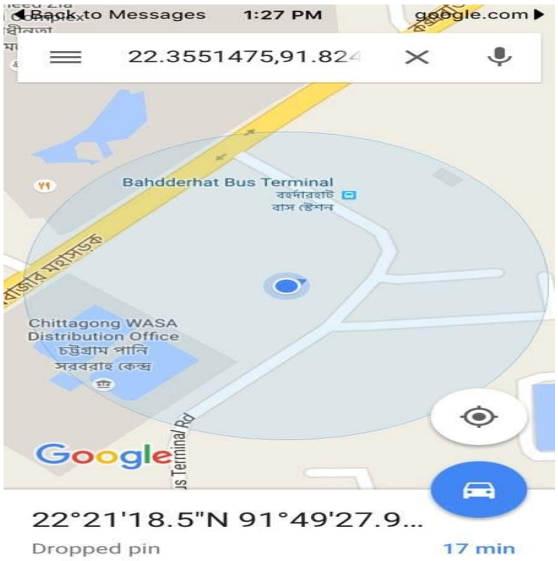


Fig. 6.Direction towards victim's location.

After sending the message to the pre defined contact list, the application sends a voice call to the top contact number. It is an automatic process of this application. When application completes the message sending task, it is automatically forward a voice call to the specific number. From this voice call the receiver can understand the current situation of the victim. The receiver also gets the message with user's GPS location, so the receiver can take action to help the user according to the situations.

Fig.7. Sending voice call to the top contact number.

III. PERFORMACE ANALYSIS

This section explains the performance analysis of the proposed system. The criterion that is considered in this section is the distance between the security device and android phone. This criterion is selected because of during the accidental situation the phone in which the application is running would be thrown away from the security device.

In the experiment it is tested whether it's getting the signal from different distances and sending the message and voice call to the correct contact numbers. The experimental result is shown in Table I.

TABLE I. PERFORMANCE ANALYSIS CONSIDERING DISTANCE

Distance (m)	Different Parameters		
	Signal	Location	Message
2	Yes	Found	Sent
5	Yes	Found	Sent
8	Yes	Found	Sent
10	Yes	Found	Sent
12	No	Not found	Not sent
14	No	Not found	Not sent

From the experimental result it is seen that the mobile android application recognized the security device signal efficiently with in the distance of 10 m. After that it is not recognized the signal from the security device. So, it is ensure that the system working perfectly within 10 m.

IV. CONCLUSION

In this paper, a generic approach is presented to reduce the attack on women and ensure the maximum security. By using this system the user can protect them self. And can get help from the relatives or known persons through the sending message and voice call automatically by pressing the security device. If this security device is used in extensively, it can be possible to reduce the violence towards the women and provide them a safe and secured life. For this system, user needs a security device made of Bluetooth device and an android mobile application. The system working properly with the distance between the security device and the mobile phone is 10 m.

Yet, the system has some limitations. Both sender and receiver need to keep the GPS on. The system not working properly with the distance between the security device and the mobile phone is more than 10 m.

REFERENCES

[1] Md. Alamgir Hossen "Measuring Gender-based violence: Results of the Violence Against Women (VAW) Survey in Bangladesh" 5th global forum on gender stistics.

[2] Dhrubajyoti Gogoi, Rupam Kumar Sharma, "Android Based Emergency Alert Button," International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, vol.2, Issue-4, March 2013.

[3] Alexandre Bartel, Jacques Klein, Yves Le Traon,Martin Monperrus, "Automatically Securing Permission-Based Software by Reducing the Attack Surface: An Application to Android" arXiv:1206.5829v2 [cs.CR] 20 Mar 2013

[4] I. S. Jacobs and C. P. Bean, "Fine particles, thin films and exchange anisotropy," in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.

[5] R. K. Sharma, A. Mohammad, H. Kalita and D. Kalita, "Android interface based GSM home security system," Issues and Challenges in Intelligent Computing Techniques (ICICT), 2014 International Conference on, Ghaziabad, 2014, pp. 196-201.

[6] Li Ma, Lei GU, Jin Wang "Research and Development of Mobile Application for Android Platform" International Journal of Multimedia and Ubiquitous Engineering Vol.9, No.4 (2014), pp.187-198

[7] Zhang Hong-Han1 , Wang Ru-Jun1 "Research on Application based on Android System" The 2nd International Conference on Computer and Applications.

[8] Jorge Zaldivar, Carlos T. Calafate, Juan Carlos Cano, Pietro Manzoni "Providing accident detection in vehicular networks through OBD-II devices and Androidbased smartphones" 5th *IEEE* workshop on user mobility and vehicular networks.

[9] Xianhua Shu, Zhenjun Du, Rong Chen "Research on Mobile Location Service Design Based on Android"

[10] Zeshan A Rajput, Samuel Mbugua, David Amadi, Viola Chepng'eno,Jason Saleem,Yaw Anokwa,Carl Hartung,Gaetano Borriello " Evaluation of an Android-based mHealth system for population surveillance in developing countries"

[11] Suhas Holla, Mahima M Katti "Android Based Miobile Application Development and its Security" International Journal of Computer Trends and Technology- vol.3,Issue-3- 2012

[12] Akanksha Singh, Arijit Pal, Bijay Rai, "GSM Based Home Automation, Safety andSecurity System Using Android Mobile Phone" International Journal of Engineering Research & Technology (IJERT) Vol. 4 Issue 05, May-2015

[13] H. Bing, "Analysis and Research of System Security Based on Android," Intelligent Computation Technology and Automation (ICICTA), 2012 Fifth International Conference on*, Zhangjiajie, Hunan,* 2012, pp. 581-584.

[14] Moon-Hee Park, Jin-Hyuk Hong and Sung-Bae Cho, "Location-Based Recommendation SytemUsing Bayesian User's Preference Model in Mobile Devices," in Proceedings of Ubiquitous Intelligence and Computing (UIC), pp. 1130 - 1139, 2007.

[15] R. Piyare and M. Tazil, "Bluetooth based home automation system usingcell phone," in IEEE 15th International Symposium on Consumer Electronics, Singapore, 2011, pp. 192 - 195.

JRanker: An Approach to Evaluate the Prestige of a Journal Using PageRank and Alexa Rank along with Impact Factor

Muhammad Kamal Hossen, Afiya Ayman
Department of Computer Science & Engineering
Chittagong University of Engineering & Technology (CUET)
Chittagong-4349, Bangladesh
mkhossen@cuet.ac.bd, ayman.fia@gmail.com

***Abstract*—For a promising academic career, nowadays it has become mandatory to publish a paper in a journal of good repute. One of the most important and possibly the least well-acknowledged aspects of the publication process is the choice of a suitable journal that is likely to help the researcher in future in various aspects. But when it comes to the assessment of 'goodness' of repute of the journal, there arrives a question how to judge this goodness. Often, the impact factor is considered an important criterion to decide the academic repute of a journal. But available impact factors present a number of shortcomings and it has become a mixed blessing nowadays. We are thus motivated to study alternative metrics that might provide a genuine measure of importance than citations alone. Such metrics exists in the form of the Google's PageRank algorithm and Alexa ranking. In the developed system, Impact Factor, Google's Page Rank and Alexa web traffic are taken into account while appraising the popularity of a journal. Considering these three factors, we developed an android based mobile application where users can find out the actual importance of a journal. The system provides more accuracy in determining the prestige of a journal than the existing systems. This system helps while comparing various journals to pick the right one for submitting one's hard work.**

***Keywords*—Journal; ISI Impact Factor; Google's PageRank; Alexa Rank; Prestige**

I. INTRODUCTION

For most research projects, the obvious goal is to have a publication in a peer-reviewed journal. The ultimate publication of one's manuscript effectively validates his/her work and can help to advance his/her academic career, attract bright students and garner funding for future studies. One of the most important aspects of the publication process is to pick a suitable journal that is likely to accept one's work and help the researcher in future in various aspects.

The most well-known measure is the ISI Impact Factor. It is considered as the most important factor to determine how good the journal is. This has caused a mad rush for publishing articles in journals which possess higher impact factors. Nowadays, it has become a dilemma for the newcomers to the field that how to search for a "Good" journal for their articles.

The most well-known measure of annually published JCR is the Impact Factor (Garfield, 1979) [1]. The impact factor is based on citations of papers published in a scientific journal. The impact factor of journal j in a given year is the average number of citations received by papers published in the previous two years of journal j. Because of its understandability and its fast availability, IF became very famous and widely used. However, one of the limitations of IF is that this ranking measure is based fundamentally on a pure counting of the in degrees of nodes in the network, and its calculation does not consider the impact of the journals in which the citations appear [2][3][4].

Impact factor may be regarded as a valuation of the citation rate of a journal's papers. Unfortunately, the JIF is now often used inappropriately, for example, in the evaluation of the quality of individual pieces of research or even the prestige of researchers.

Google's PageRank is a popular webpage ranking algorithm; it views a hyperlink as a recommendation [5]. The computation of the scores of webpages is based on a combination of the number of hyperlinks that point to the page and the status of pages that the hyperlinks originate from. However, similar to other recommendation systems such as journal citations, the status of the recommender also bears importance. Google's PageRank takes into account the status of the recommender. A page can be considered as important if it is pointed to by other important pages.

Alexa ranking is also a very popular web ranking method which works the same way as Google's PageRank. This ranking system is set by alexa.com that basically inspects and publishes the frequency of visits to various Web sites in various categories. The Alexa [6] is an authority site to evaluate the common webs, and the Alexa gives each website on the internet a ranking number to show its popularity. While ranking webs, the Alexa takes lots of properties, such as click rate, In-links and Out-links; they reveal the information performance from another aspect, which is different from the traditional evaluation in response time, reliability and so on.

There are certain benchmarks that go along with the rankings. Ideally, any website wants high PageRank and low

Alexa Rank. A PageRank of 5 is good. An Alexa Rank of anything under 100 thousand is great.

We are proposing a way to measure the influence of a journal by taking into account some other factors rather than only impact factor. The system considers two ranking methods along with impact factor while evaluating a journal's importance.

The main objective of this system is to provide a truer ranking for a journal with the help of three factors:

- Journal Impact Factor
- Google's Page Rank
- Alexa Web Rank

Impact factor gives an idea about the citation rate of any journal. While PageRank measures a website's relative strength and affects Search rank, Alexa is more like facts and figures based on web traffic. It shows how many people are visiting a site.

So it is reasonable that we adopt PageRank and Alexa for journal ranking. PageRank and Alexa, as an alternative to the ISI's impact factor, undoubtedly represent substantial advantages.

II. FACTORS FOR EVALUATING JOURNAL'S IMPORTANCE

A. The ISI Impact Factor

The impact factor is based on citations of papers published in a journal. Institute for Scientific Information (ISI) has published them since 1961. The higher the value of impact factor, the higher the scientific esteem of the journal. The founder of the Journal Impact Factor (JIF) Eugene Garfield had originally designed it as a means to help choose journals.

The impact factor is defined as a ratio of two elements: The total number of citations in the current year to any article published in a journal during a given time period and the total number of published citable articles in that particular journal within that time period. This timeframe is defined as two years by ISI.

The impact factor of journal A in a particular year Y is calculated using the formula given below:

$$IF_A = [\text{All citations in Y to articles in A during } (Y\text{-}1) + (Y\text{-}2)] \,/\, [\text{All citable articles in A during } (Y\text{-}1) + (Y\text{-}2)] \qquad (1)$$

We observe that the ISI impact factor is based fundamentally on a pure counting of the in-degrees of nodes in the network.

B. Google's PageRank

The inventor of Google's PageRank algorithm is the same as its founder; Larry Page and Sergey Brin. The scores of web pages are computed based on a combination of the number of hyperlinks that point to the page and the status of other pages from which the hyperlinks originated[6].

The original PageRank algorithm is given below:

$$PR(A) = (1-d) + d\,(PR(T1)/C(T1) + \ldots\ldots + PR(Tn)/C(Tn))$$

Where-

- PR(A) = PageRank of page A
- PR(Ti) = PageRank of pages Ti which link to page A
- C(Ti) = Number of outbound links on page Ti
- d = Damping factor having value between 0 and 1, usually set to 0.85.

Every few months Google calculates pages PRs (PR update). During this update, all pages are assigned a new PR by Google and users will have this PR until a new update is done. New sites that are just launched will have a PR of 0 until a new PR update is done by Google.

C. Alexa Global Ranking

Alexa's Traffic Ranks are based on the traffic data provided by users in the global data panel of Alexa over a 3 month period. Alexa updates Traffic Ranks daily. A combined measure of Unique Visitors and Pageviews determine a site's ranking. 'Unique Visitors' is a variable which is determined by the number of unique Alexa users who visit a site on a given day. 'Pageviews' is indicated by the total number of Alexa user URL requests for a webpage. Multiple requests from the same user for the same URL on the same day are counted as a single Pageviews. The site which possesses the maximum combination of Unique Visitors and Pageviews is ranked #1.

The smaller ranking number in Alexa can show higher popularity in users' experience, which reveals higher performance in information providing evaluation. Therefore, we import the Alexa ranking result to assist evaluating the performance of a journal in our approach.

D. Our Proposed JRanker Method

For the proposed JRanker method, we considered all the three factors mentioned above.

As Thomson Reuter Impact Factor is authentic, so we chose not to alter it in any way. But in case of Google's page rank and Alexa we did some modification. We converted it to a scale of 5 to combine these two factors together.

After the input is given, the system retrieves the corresponding Uniform Resource Locator (URL) from the journal database. The retrieved URL is fed into the PR fetching function as an input. This PR fetching function returns the page rank of the website of the journal. It sets agent and host to request Google toolbar. After the response of agent, hash map is implemented on the received data to extract the desired page rank.

Page rank is calculated on the scale of 10. As the final JRanker score is calculated on the scale of 10, we assigned both factors (page rank and Alexa rank) a measurement on the scale of 5. So, for the proposed system, the rank is converted to a scale of 5 by multiplying the result by 0.5 and F_b is obtained.

In case of Alexa ranking system, the lower rank indicates the higher impact. But for the proposed JRanker system, this ranking is reversed according to the following algorithm.

```
Let Alexa_Global_Rank = G.
If G<=100000
    {
        Fraction = G/10000;
        Reversed_Rank= 10 - Fraction;
    }
  Else
   Reversed_Rank = 0.1
```

If the rank is lower than 100000 then it is being reversed by subtracting a fractional value from 10. The fractional value is obtained by dividing G by 10000. This gives us a score up to 4 decimal places on the scale of 10. In case of global rank higher than 1 lac the reversed rank is set to 0.1. Because an Alexa rank that is high doesn't really bear much impact anyway. So, ranks higher than 100000 are set to 0.1 in reversed ranking.

The retrieved URL from journal database is sent to data.alexa.com. When it responses, the Alexa global ranking is fetched by segmenting the received data.

The received global rank is then reversed by using the formula explained above. For the proposed system, the rank is then converted to a scale of 5 by multiplying the result by 0.5 and F_c is obtained.

For final score, we combined our F_b and F_c which gives us the desired JRanker score on a scale of 10.

III. System Description

A. System Architecture

The development of this system requires two steps. First, we retrieve the impact factor of a journal and the corresponding website address of it. Then we obtain the page rank of the journal and Alexa global ranking for the website of the journal. Then these two results are combined together to produce a truer measurement of the popularity of the journal.

In the first step, the name/title of the journal will be given as input. Then the processing module will request 2014 impact factor of the journal along with PR data and Alexa data. We couldn't use 2015 impact factors as they will not be published until June, 2016.

In case of Alexa ranking, the obtained value will be reversed to get a similar result on the scale of 10. After reversing Alexa global rank and obtaining PR, they will be converted to a scale of 5 in order to generate a score on the scale of 10.

The system also provides the user with the impact factor list of 2014 published by Thomson ISI for almost 800 journals.

There is also a side feature in the system which is called conference calendar. This feature provides the upcoming conference dates, venues and submission deadlines of the conferences to be held in Bangladesh.

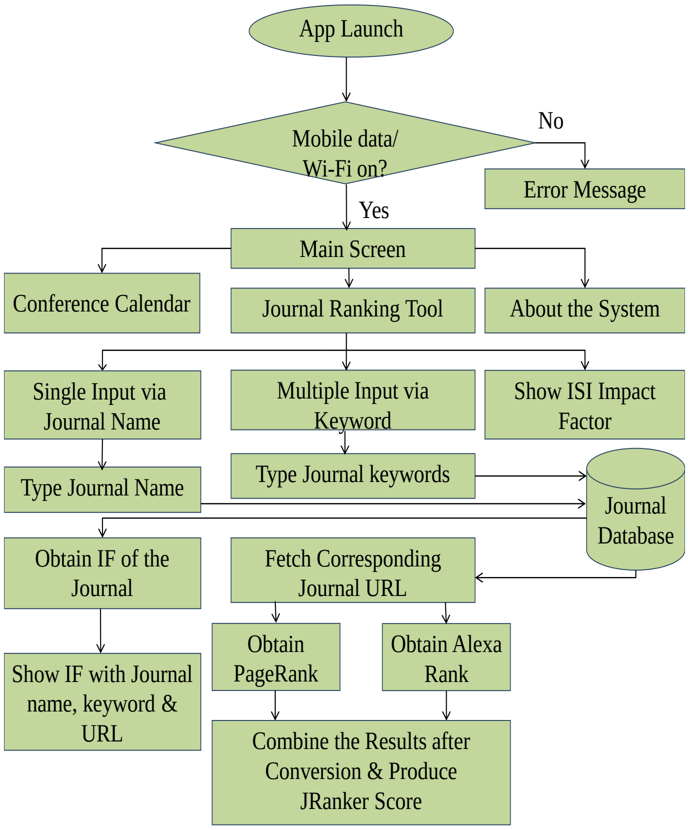


Fig. 1. System Architecture.

B. Mobile Application

This application is developed on android platform. The front end is the visual part of the application that the user interacts with, and the back end, which contains all the code that drives the system.

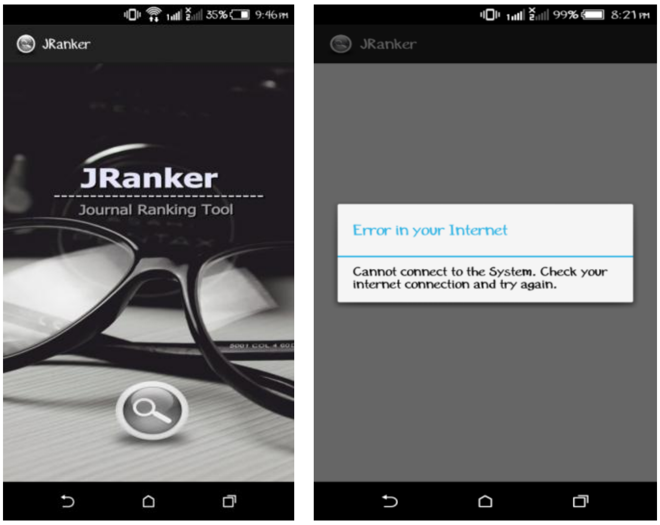
As the application is a Dynamic Application, it needs to access a server in order to properly function, i.e. it needs a data connection either through Wi-Fi or phone's carrier.

When the network is established, the application loads the content from the server after the splash screen. If the connection is failed somehow, then it shows the error message just after the splash screen which is manifested in Fig. 2(a) and Fig. 2(b).

There are three buttons/options on the home page. They are-

- JRanker
- Conference Calendar
- About System

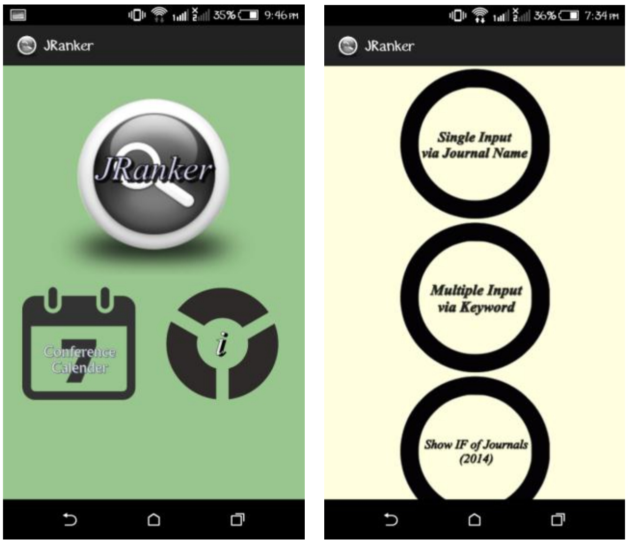
The first two buttons provide the service to the users while the third button provides a brief knowledge about the system. The main page is shown in Fig. 2(a).



(a) (b)

Fig. 2. (a) Splash Screen, (b) Error Message.

The main purpose of this system is to measure the popularity of a journal with the help of Impact Factor and two web ranking algorithm. The first button leads the user to the next page which contains three options. The system provides these to the user. They are single input via journal name, multiple inputs via journal keyword and the total 2014 impact factor list of journals published by ISI in June, 2015. These options are shown in Fig. 3(b).



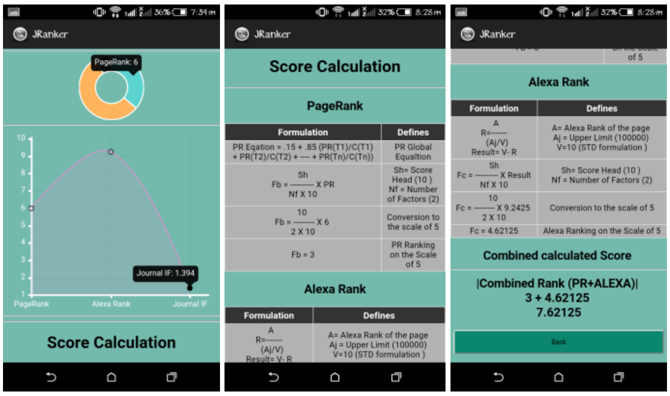
(a) (b)

Fig. 3. (a) Main Screen, (b) Three Options for Checking Popularity of Journals.

In “Single Input via Journal Name” option, the user needs to type the journal name accurately in the given box. If the exact suggestion appears, the user needs to select it to see the result. The user needs to keep typing the journal name until he/she gets the desired journal name. Fig. 4(a) shows the suggestions which will appear if ‘journal’ is typed in the given box as input.

(a) (b)

Fig. 4. (a) Suggestions, (b) Output for JACM with Table & Bar Chart.



(a) (b) (c)

Fig. 5. (a) Output for JACM with Donut Chart & Line Graph, (b) Score calculation of PageRank, (c) Score Calculation of Alexa & Final Score.

After selecting one suggestion, the system then connects to the database to retrieve 2014 impact factor and the corresponding URL stored in the database. The system calls the Google toolbar API and Alexa API and feeds the API calling functions the given URL as input. Upon receiving the data, the system implements hash function on them to extract the desired value. The values are showed as output with the help of bar chart, donut chart and line graph. The output for Journal of the ACM is shown in Fig. 4(b) and Fig. 5(a).

The Score Calculation table gives the user a thorough idea of the three ranking system used in the application. In Fig. 5(b) we can see the score calculation of PageRank. The global PR equation of calculating Google’s page rank is given. Then we can see how the results are converted to a scale of 5. To provide a clear understanding, we showed the calculation precisely.

The table in Fig. 5(c) shows the score calculation of Alexa rank. Alexa rank is reversed at first following the formulas described in section 3.4.3. After reversing the rank, it is then converted to a scale of 5 for the developed system.

We combined the PR and Alexa global rank together to produce a score based on a scale of 10. Considering the two factors separately might confuse the user, that's why we combined them to produce the final score.

In "Multiple Inputs via Journal keywords" section we introduced a different keyword system. The user needs to take the initials of the words from the journal name except for articles and prepositions. Such as, for journal name ACM Transactions on Architecture and Code Optimization, the keyword will be ACMTACO. For multiple inputs, the user needs to write the keywords using coma to separate them. The bulk input is demonstrated in Fig. 6(a).

Four keywords have been given as input. The outputs for these four journals are shown in Fig. 6(b), Fig. 6(c), Fig. 7(a) and Fig. 7(b).

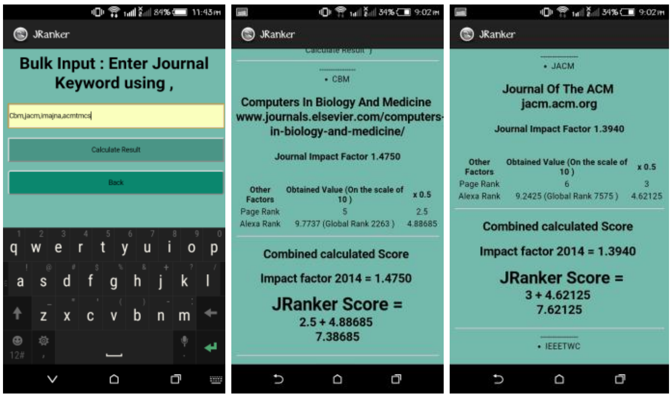


Fig. 6. (a) Bulk Input, (b) Output for CBM, (c) Output for JACM.

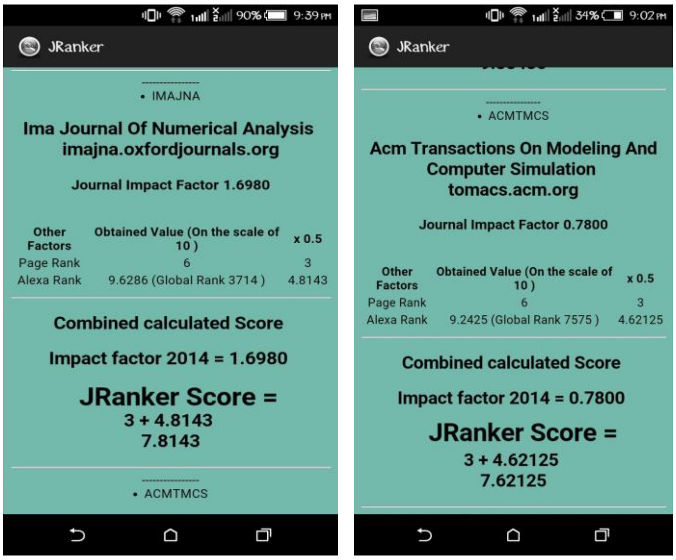


Fig. 7. (a) Output for IMJNA (b) Output for ACMTMCS.

The system can currently measure prestige or popularity of almost 1500 journals in the fields of CSE, EEE, Science and Mathematics etc. In this application, we provide the latest impact factor list published by ISI for these journals. A few screenshots of the total list are given in Fig. 8.

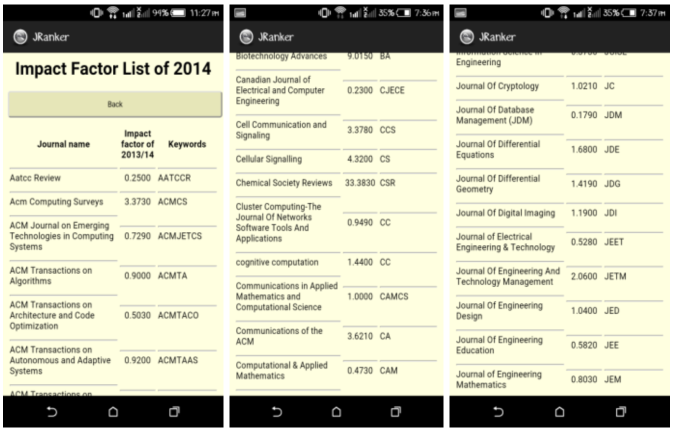


Fig. 8. ISI Impact Factor List (2014) of Journals.

We created a database in the server to store corresponding website address for a given journal. When the application launches, the system connects with the database for further operation.

IV. Experimental Results & Evaluation

We applied our ranking system to measure the real impact of various journals. For evaluation part, the journals we used are Journal of the ACM (JACM), IEEE Computational Intelligence Magazine (IEEECIM), International Journal of Computer & Electronics Research (IJCER), International Journal of Scientific Research and Development (IJSRD). The first two journals are prominent and the latter two are known to be predatory journals.

TABLE I. Difference Among Three Factors for Each Journal

Journal Name	Impact Factor	Page Rank	Alexa Rank
JACM	1.394	6	9.2382 (Global Rank 7618)
IEEE CIM	2.571	7	9.768 (Global Rank 2320)
IJCER	2.4025	5	0.1 (Global Rank 11351408)
IJSRD	2.3898	0	0.1 (Global Rank 657323)

The obtained values from the system are shown in Table I. The impact factors of these four journals are almost similar. The latter two journals used impact factor values from fake provider in their own websites in order to make publicity. If we only consider IF to evaluate the prestige of these journals, we cannot distinguish between prominent and predatory journals.

TABLE II. CONVERTED VALUES & FINAL SCORE

Journal Name	Converted PageRank (PR value x 0.5)	Converted ALEXA Ranking (Converted value x 0.5)	Final Score (PR + Alexa)
JACM	3	4.6191	7.6191
IEEE CIM	3.5	4.884	8.384
IJCER	2.5	0.05	2.55
IJSRD	0	0.05	0.05

Now considering Pagerank and Alexa rank of the websites of these journals, we can get a precise idea about the popularity of these four journals. The final score shows much discrimination in the popularity of these journals than impact factor. Table II shows the converted values and final score produced by our system.

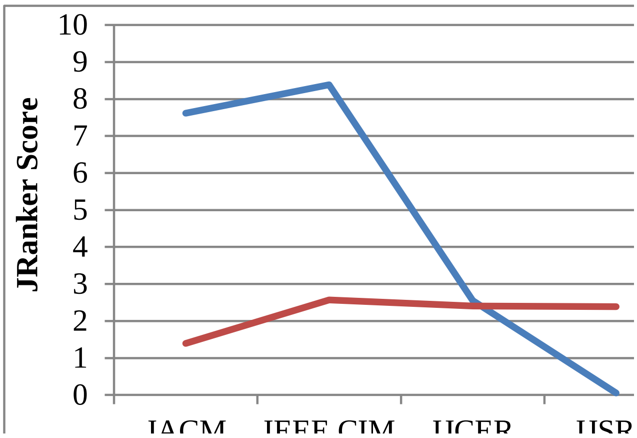


Fig. 9. Comparison between JACM, IEEE CIM, IJCER and IJSRD based on final score.

In Fig. 9, the red line shows the impact factors of the four journals and the blue line shows the final score produced by the developed system. By analyzing Fig. 9, we can see that though impact factor doesn't distinguish among the popularity of these four journals, the single score produced by the developed system based on Pagerank and Alexa rank clearly discriminate among these journals. The difference between the prominent journal and predatory journal is clearly visible based on the score produced by the developed system.

The application isn't uploaded to Google play store yet. But anyone can download the apk file from the link given below: *https://drive.google.com/open?id=0Bzcsx8v82dQSbFpQRFJTaG1jRGs*

V. CONCLUSION

Choosing a journal with better reputation is very important before submitting one's research. Often, the impact factor is considered an important criterion to decide the academic repute of a journal though it is not advised. We developed a system which implements two web ranking algorithms to evaluate the popularity of a journal; Google's PageRank and Alexa global ranking. The developed system evaluates a journal's prestige precisely and far more accurately than Impact factor alone. The comparison section shows clearly how effective this system can be while appraising a journal and it discriminates among journals effectively. This application works very smoothly and without any delay. It will help a lot of people who are engaged in research work. Some future direction for this application can be the following: other webpage ranking methods can be considered for this evaluation. Again, implementation of this system can be for different platforms.

REFERENCES

[1] E. Garfield, "Citation Indexing: Its Theory and Application in Science, Technology, and Humanities," John Wiley and Sons, New York, 1979.

[2] G. Buela-Casal, "Assessing the Quality of Articles and Scientific Journals: Proposal for Weighted Impact Factor," Psychology in Spain, 8(1), pp. 60-76,2004

[3] M.Amin and M. Mabe, "Impact factor: use and abuse," Perspectives in Publishing, pp. 1-6, 2000.

[4] J. Bollen, M. A. Rodriguez, H. V. de Sompel, Journal Status, Last day of access (August 22, 2015). [Online]. Available: http://arxiv.org/abs/cs.DL/0601030

[5] S. Brin and L. Page, "The anatomy of a large-scale hypertextual web search engine," Computer Networks and ISDN Systems, 30(1–7), pp. 107–117, 1998.

[6] Alexa. Last day of access (August 22, 2015). [Online]. Available: http://www.alexa.com/

[7] E. Garfield, "The history and meaning of the journal impact factor," Jama 295.1, pp. 90-93, 2006.

[8] D. Munger, "Can journals game the impact factor system?," Last day of access (August 22,2015). [Online]. Available: http://scienceblogs. com/cognitivedaily /2006/06/08/ can-journals-game-the-impact-f/

[9] S. Cheng, Pan YunTao, Yuan JunPeng, Guo Hong, Yu ZhengLu, and Hu ZhiYu, "PageRank, HITS and impact factor for journal ranking," in Computer Science and Information Engineering, WRI World Congress on, vol. 6, pp. 285-290, IEEE, 2009.

[10] S. Prabha, K. Duraiswamy, and J. Indhumathi, "Comparative Analysis of Different Page Ranking Algorithms," Int. Journal of Computer, Electrical, Automation, Control and Information Engineering vol. 8, no. 8, 2014.

[11] R.S. Shekhawat, P. Setia, and D. Awasthi, "Higher Impact Factor: Better Journal? Not a Necessity," Journal of Indian Academy of Forensic Medicine vol. 36.2, pp. 199-202, 2014.

[12] P. V. Kamat, G. C. Schatz, "Journal Impact Factor and the Real Impact of Your Paper," The journal of physical chemistry letters, vol. 6.15, pp. 3074-3075, 2015.

[13] P. Devi, A. Gupta, and A. Dixit, "Comparative Study of HITS and PageRank Link based Ranking Algorithms" International Journal of Advanced Research in Computer and Communication Engineering, 3.2, 2014.

A System to Ensure Privacy for Android Users

Pranab Kumar Dhar*, Sulogna Chowdhury, and Lamia Alam
Department of CSE, Chittagong University of Engineering and Technology (CUET), Chittagong-4349, Bangladesh
Email: pranabdhar81@gmail.com, konthichy@gmail.com, and lamiacse09@gmail.com

***Abstract*— Now-a-days, android based smart phones are very popular, because they offer novel ways of communication, information search, sharing, and entertainment. But still there are some issues related to memory usage, privacy and security which need to be considered. Many unused cache, cookies and junk files are generated by various applications which take unnecessary space and slow down the system. Besides, getting call from unwanted number with an intention to harass and disturb people is a common problem. Moreover, users today are also concerned about keeping their personal files protected from unauthorized access. To eliminate all these problems, in this paper, we propose a system to ensure privacy for android users. It has four specialized features in one application such as call blocking, file locking, application manager and cache cleaning to make the operating system more user-friendly. Experimental results indicate that around 76% user found this application very useful.**

Keywords—Smartphone; android; cache cleaning; privacy; call block; file locking;

I. INTRODUCTION

Android is a Linux based operating system made for the touch screen smart phone device. It is an open source platform, customized by many major companies. People are getting more attracted to this platform because of the vast functionality and easy access of the applications. Besides, it provides the facilities of faster and easier browsing.

Android device provides a lot of functionality. Calling to the desired person is one of them. But it has drawbacks also. Sometimes repeated calls from unwanted caller with an intention to harass can be annoying and threatening. In addition to that, as android phone can be treated as a computer in respect of its multipurpose function. These files contain important as well as confidential elements such as personal information. Another most common function of android phone is running various types of application for the betterment of the device. These applications should be managed carefully. Some applications need huge space in the memory. When the internal memory becomes almost full, the device becomes slow. So, there should be an option in the device to manage the files properly and to uninstall any file when necessary.

To run different kind of applications and other files cache memory is needed. A 'cache' is a temporary place where developers store data of a particular website or application software (apps). When the user accesses that specific website or apps again the data from the cache files are used to load the pages instead of downloading it again, because accessing to the network is slow and expensive. The main purpose of caches is to improve browsing experience by boosting up the page loading speed. However, along with all the speed advantages they also have some disadvantages that are enough to get rid of them. These files are stored in user's phone memory which means that they are definitely taking some space. Furthermore, if they are not deleted from time to time they can also lead to slowing down the phone's speed or even stop the web pages and apps from working properly. There are many advantages of clearing cache files that easily outweigh the advantages of keeping them. To eliminate all these problems, in this paper we propose a system to ensure privacy for android users. The system has four specialized features in one application. These are call blocking, file locking, App manager and cache cleaning, respectively. By call blocking feature the user gets the opportunity to block several numbers or restrict them to call him/her. By this feature the user gets rid of unwanted and harmful calls. To boost up the speed of the phone the user can clean the cache frequently by using this system. The user also can know about the previously used cache memory by this application. To hide any confidential or important files from other users, he/she can lock any file he/she wants and can restrict the files from further access to other persons. By the app manager option, the user can manage an application running in his phone in a very suitable way. He/she can run any application very quickly. On the other hand, if any application is no longer needed, the user can uninstall it by using this system. All the four specialized features are implemented in the proposed system effectively. To the best of our knowledge, all these four features are incorporated in one application software for the first time. The android users can access their device more effectively by using this proposed application.

II. RELATED WORK

Memory usage, data protection and user privacy are challenging topics because of the diversity of potential issues and threats. There has not been any great advancement in the field of security and privacy for android mobile user. Finley *et al.* [1] proposed a system of caching the android mobile phone. This system presents a dynamic cache cleaner that aggressively pushes out unused cache data. But it can be done by manual process and it only manages the cache file. In [2], authors described a system for multi-user environment which prevents unauthorized access of user. This system is specified for three users at a time and the primary user has to select

*Corresponding author

other user's access option. Moreover, there is no specification of temporary file management of multiple users. An encryption file system is proposed by Wang *et al.* [3] for protecting personal identification information of removable and persistent storage to secure user's privacy. It is applicable only for image file and music file. The authors of [4] presented a static analysis framework for finding potential leaks of sensitive information automatically in android applications. This system not only secures the selected private information but also notifies if any information is leaked. There is another application available named 'Clean Master' [5] provided by Google. This application software has the feature such as cleaning junk file manually and memory booster to speed up the memory. But android phone has its own application management system. Therefore, it does not need additional system for memory boosting. Yajinet *et al.* [6], showed a comparative analysis of the vulnerability of security for using various apps in android. They investigated how private data are leaking everyday and user's information is getting public. A user friendly app named secure gallery [7] which gives the opportunity to hide the user's private pictures and videos.

In this paper, we propose a frame work that has four specialized features in one application such as call blocking, file locking, application manager and cache cleaning to make the operating system more user-friendly.

The main advantages of proposed system over previous work are as follows:

- All the four features like file locking, cache cleaning, call blocking and application management can be incorporated within a single system.
- Any type of file can be locked by the file locking system. In previous work [3], only music and image files can be locked.
- Multiple call blocking can be done using this system. System cloud is used by which blocked numbers can be updated and deleted easily.
- Cache cleaning system is automatic. In previous work [5], the cache can be cleared manually.

III. Proposed System Architecture

The schematic representation of our proposed system is illustrated in Fig. 1. The whole system consists of four major modules. These are File Lock, Cache Clear, Call Block and Application Manager. Detailed processing of these modules is given in the following subsection.

A. File Lock

File lock is one of the major features of our proposed system. By using this feature particular folder or file in the phone can be locked. A password can be used to lock the folder. After loading the file there are two options. If user wants to lock the file, he needs to give password. The system works by the following way. At first, the system gets the file to be locked. When the password is given, the file name is changed by adding an extension to the file name. The password is saved to the shared preference. When unlocking, the system checks for the correct password in shared preference. If it is matched, the extension is removed and the file will open. Otherwise it will show invalid password. Fig. 2 shows the flow chart for file lock.

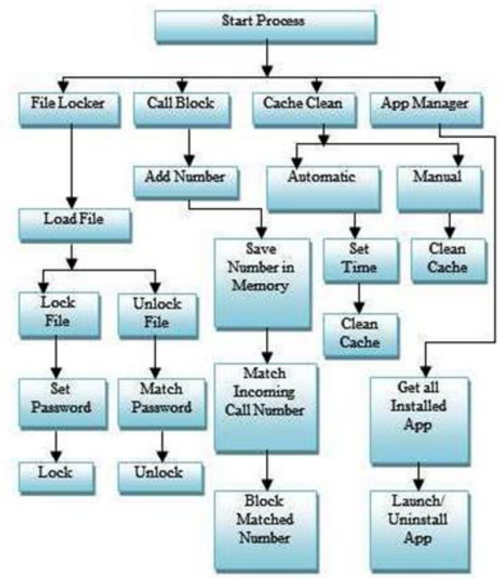


Fig. 1. Proposed framework

B. Cache Clear

Caching is a mechanism for transparently storing data in such a way that the future requests can be accessed more efficiently. The performance of the application is directly affected by the size of cache memory. If the cache memory is not clean regularly, the phone may become slow. In this application, there is an option of cache cleaning. At first, the screen shows the total used cache. Then the cache cleaning option arises. Here, the cache cleaner works in a particular way. At first it accesses to the cache directory and check for the cache file. Then the size of the file is calculated (1 cache = Total cache/1024) which shows the total cache memory in megabyte. After that, the detected file is deleted. Therefore, it can clean cache automatically while other systems clean manually. Fig. 3 shows the flow chart of cache cleaning process.

C. Call Block

Call block is an important feature of this application. After the launching of application a particular option named 'Call Block' arrives. In this option, several numbers can be blocked or restricted. When the user gives the number, it will save it in memory. When there is a call from the blocked number, it will verify the number and disconnect the call. The user will not be notified about this call. To keep aloof from unwanted calls this feature is very much effective. Fig. 4 shows the flow chart of call block.

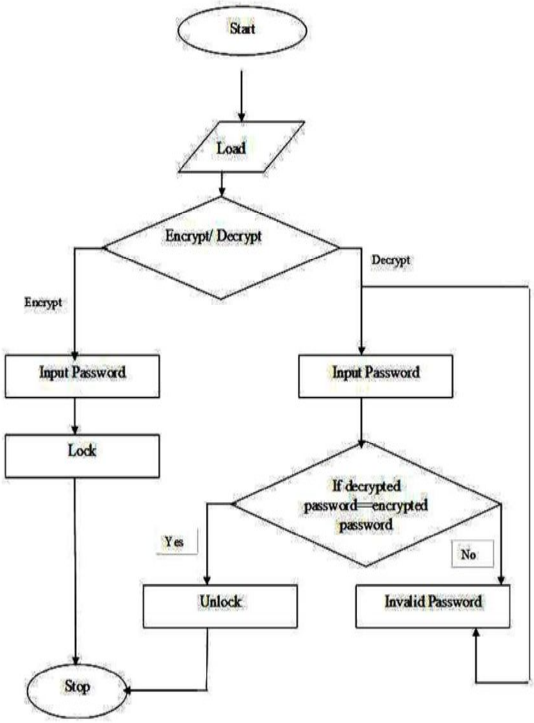


Fig. 2. Flow chart of file lock

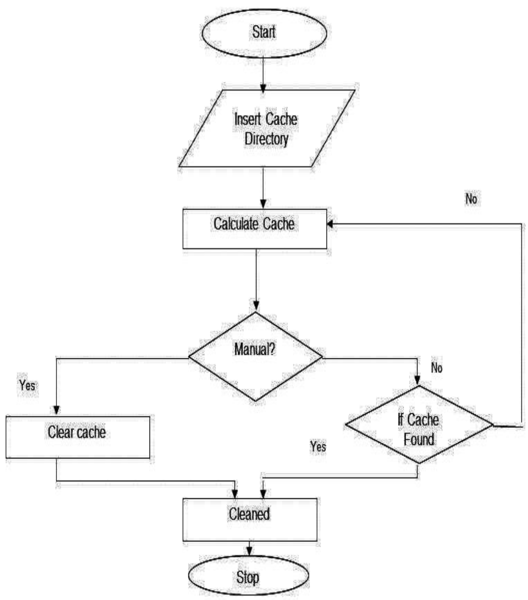
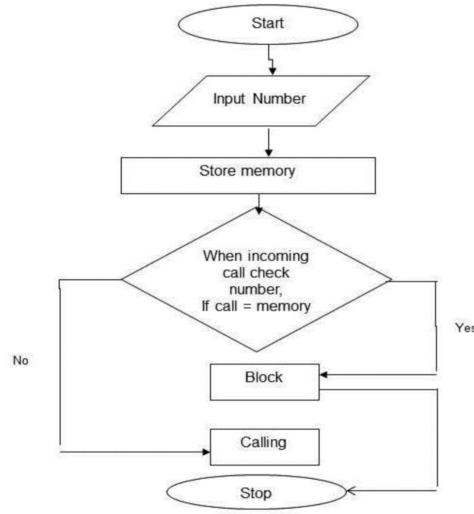


Fig. 3. Flow chart of cache clean



To block multiple phone numbers, the user has to login to the system cloud. In the system cloud the user can add more numbers to block. After that Json array is updated and http requires Json parse. After creating SD folder, the system will search number and save the numbers in android shared preference database. Fig. 5 shows the flow chart.

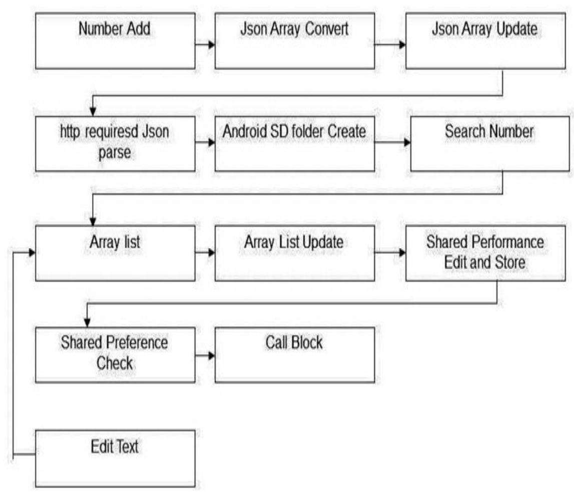


Fig. 5. Flow chart of multiple call block

D. App Manager

The last feature of this application is application manager. After starting the process, it will get access to the entire installed app in the device. Then the user may select whether to launch any app or uninstall it. By this feature particular application can be used. On the other hand, any application can be uninstalled. Fig. 6 shows the flow chart of app manager.

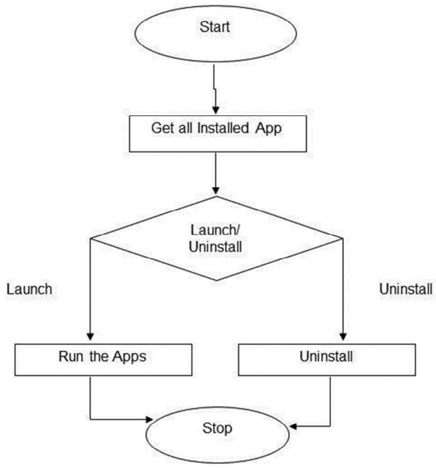


Fig. 6. Flow chart of app manager

IV. IMPLEMETATION

The aim of this work is to build a common platform to ensure all the specified features. Fig. 7 provides screen shot of various user interfaces for our system. In order to develop this system JAVA, PHP, MySQL, Apache Server, NetBeans, Android Emulator, Symfony, JSON etc. have been used. Using these technologies we developed a system that can perform the following tasks:

- Ensure unique and multiple file protection system
- Speeds up the system
- Ensure to clean unused cache files, cookies and junk files
- Blocks specific phone calls
- Easy access to all installed apps on device.

Fig. 7 provides screen shot of various user interfaces for our system.

V. USER RESPONSE AND FEEDBACK

In order to measure the acceptability and usability of the system, we conducted experiment and then compared the proposed system with other existing application.

We have collected data from the people who are using android for so many days. We asked them about their cache cleaning system, system speed, about unwanted phone calls, if they use any system for preventing unwanted calls, and what process they follow to protect their private files. Most of them use several systems for doing these and some are not much concerned about their privacy. We asked the user about their expectations of using android system more efficiently. Based on the conversation, we summarized the basic requirements of the user which are as follows:

- Automated cache cleaning system
- More speedy operating system
- Easier File management system
- To get rid of unwanted phone calls
- Reduce the complexity of using different apps
- A smart way to maintain all of these features properly.

We then asked participants to fill out a questionnaire for each condition. The measurement was done using the grades: (1) Yes, (2) No, and (3) Not needed. The questionnaires had the following items:

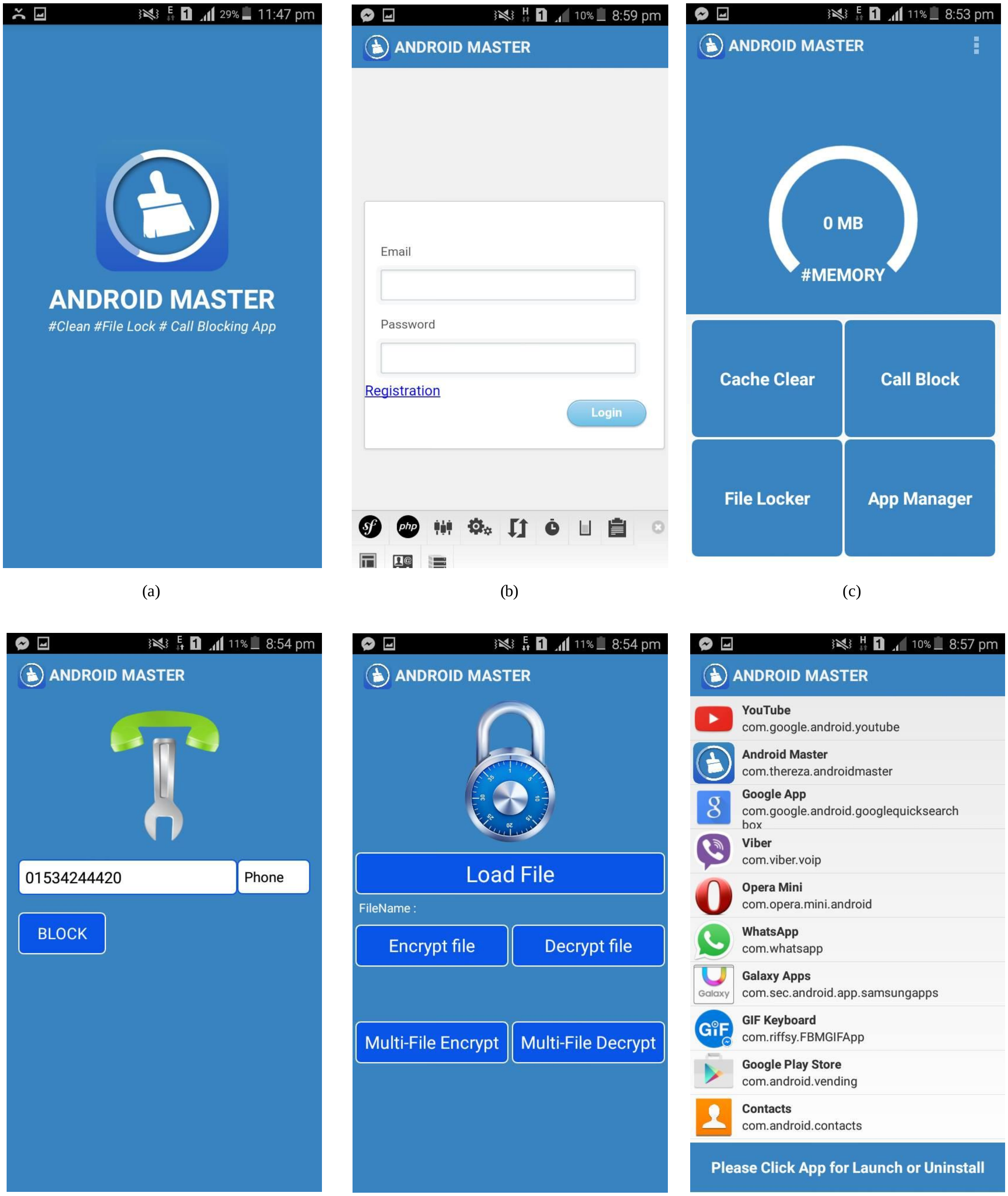
- Does the cache cleaning system make your device slow?
- Does the application takes much space in memory?
- Can you manage multiple numbers properly with call blocking system?
- Are you satisfied with the call blocking system?
- Does the system can protect multiple file/content?
- Do you have easier way to manage your entire installed apps on your device?
- Are you satisfied with content protecting system?

Fig. 8 illustrates the assessment of questionnaires for the proposed system and the existing systems. In Fig. 8, x-axis represents the sets of questions and y-axis represents the ratings of the people on the basis of the questions. The user answers are mostly negative in case of existing system whereas user response is mostly positive in case of proposed system. This means that most of the users were satisfied with the system that fulfills their demand.

In order to evaluate our proposed system, we have also collected feedback from 100 people and asked them to rate this application using three grades: (1) Very useful, (2) Helpful and (3) Need Improvement. Our findings regarding privacy settings and the evaluation results statistics based on the public opinion are presented in Fig. 9. Almost 76% of reviewers replied on this application as a very useful one. 18% people justified the project as helpful to them and the rest 6% people suggested for improvement. The evaluation of the statistics based on public opinion eventually gives us the concept that the application is applicable and acceptable to the android phone users.

VI. CONCLUSION

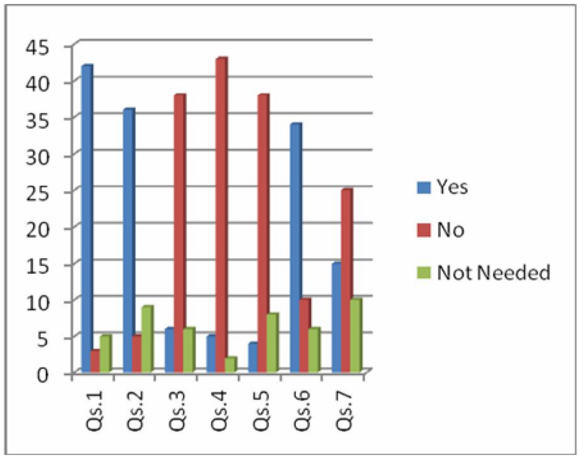
In this paper, we have presented an application which is a combination of various features such as call blocking; file locking, App manager and cache cleaning. This application will cost almost nothing to the users. It is a free version and can be easily downloaded or collected. To maintain the privacy of the users, it can be a great solution. This application restricts the device from any kinds of cyber crime. Leakage of personal data can also be prevented. The users can block any unwanted call. All the features can be accessed by using this single application. The android users will be very much benefited if they use this application. Our evaluation of the statistics based on public opinion demonstrates that the application is applicable and acceptable to the android phone users. In future, we will develop this application in cross platform. Moreover, the call block system, spam and malware detection will be improved.



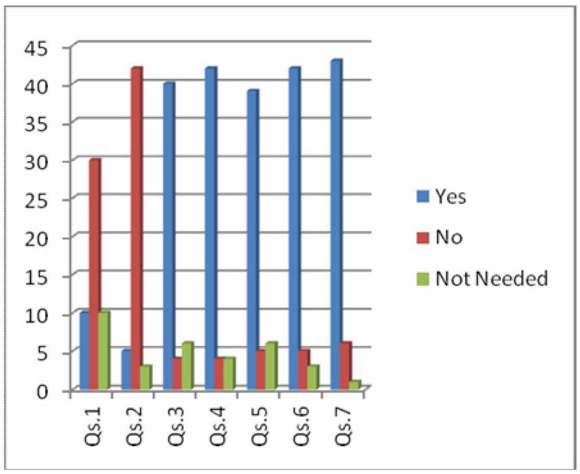
(a) (b) (c)

(d) (e) (f)

Fig. 7. (a) Starting window (b) Login window (c) Homepage (d) Call block (e) File lock (f) App manager screen



(a)



(b)

Fig. 8. Results of questionnaire assessment (a) Existing System and (b) Proposed System

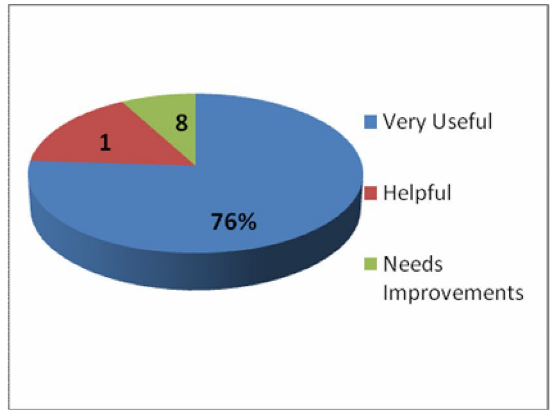


Fig. 9. User Review of the Application

REFERENCES

[1] S. Finley and X. Du, "Dynamic cache cleaning on android," *in Proc. IEEE International Conference on Communications (ICC-2013)*, pp. 6143 – 6147, 2013.

[2] W. Du, P. Rattazi, H. Hao, Y. Aafer, A. Ahlawat, and Y. Wang, "A systematic security evaluation of android's multi-user Framework," *in Proc. 3rd Workshop of Mobile Security Technologies (MoST)*, pp. 1-10, San Jose, CA, USA, 2014.

[3] Z. Wang, R. Murmuria, and A. Stavrou, "Implementing and Optimizing encryption file system on android," *in Proc. IEEE 13th International Conference on Mobile Data Management (MDM)*, pp. 52-62, 2012.

[4] C. Gibler, J. Crussel, and H. Chen, "AndroidLeaks: Automatically Detecting Potential Privacy Leaks In Android Applications on a Large Scale," *in Proc. 5th International Conference on Trust and Trustworthy Computing (TRUST)*, pp. 291-307, 2012.

[5] Clean Master (Speed Booster) (Version 5.9.6), [Mobile Application Software]. https://play.google.com/store/apps, Last day of access (April 9, 2015).

[6] Y.Zhou, and X. Jiang,"Detecting Passive Content Leaks and pollution in android mobile," *in* Proc. *20th Network and Distributed System Security Symposium (NDSS)*, pp. 2013.

[7] Secure Gallery (version 3.3.0), [Mobile Application Software], Available: https://play.google.com/store/apps, Last day of access (April 18, 2015).

A location based smartphone application to rent private vehicles at real time

Faisal Ahmed
Dept. of Computer Science & Engineering
University of Chittagong
Chittagong - 4331, Bangladesh
Email: faisal.cse91@gmail.com

Fatema-Tuj-Johora
Dept. of Computer Science & Engineering
University of Chittagong
Chittagong - 4331, Bangladesh
Email: inna.johora@gmail.com

Mohammad Sanaullah Chowdhury
Dept. of Computer Science & Engineering
University of Chittagong
Chittagong - 4331, Bangladesh
Email: sana1691@gmail.com

***Abstract*—Although present era is the age of the fastest communication and travelling, some aspects are still there which demand our noticing and adding comfort to. The unavailability and poor quality of public transports like bus, subway or train have made locating private vehicles such an alarming event. People are frequently a victim of harassment to locate vehicles at the right time and at the actual rent. To make an end of all troubles in renting vehicle, we are proposing a location based smart system to facilitate the passengers as well as the drivers. This paper outlines a client server based smartphone application having easy access for the passengers to locate their nearby private vehicles, to invite drivers for a rent and to bargain on the rate on private chat or voice call. Drivers will also get convenience in searching passengers who are awaiting them rather wandering randomly. To ensure safety, riders can rate drivers instantly depending on driving skill and easily see their overall rating. Reservation is attainable through this application. The recommended system takes advantage of Java programming language using android operating system for mobile client, PHP as web server, MySQL as repository, GPS(Global Positioning System) as location provider and Google map service for showing location on mobile client in conjunction with the street view between the passengers and the drivers.Emperical result shows that the average location accuracy of the application is about couple of meters.**

Keywords: GPS, **GSM**, **LAMP**, **MySQL**.

I. Introduction

Searching out private vehicles or helplessly praying for one to miraculously pass by our location has become a common event in our everyday life. It is nearly impossible to rent a vehicle at the right time and at the actual rent when people are travelling to and from work at peak hour. Being absent in workplace at proper time and causing death of emergency patients due to unavailability of vehicles at the proper time are emerging as a serious issue. Sometimes one has to go the opposite direction in search for a private vehicle and wait a long time in the street, which is tedious. Further one has to ask same and several questions regarding rent by going driver to driver which gives rise of untoward incidents every so often.

As a very little work in reality has been actually done in this aspect, we are proposing a location based real time smart private vehicles searching system that can solve such type of problems without any annoying hold music and enigmatic whereabouts of the vehicles.

The latest annual report from ERICSSON [1] reveals that the number of smartphone users will reach a giant 6.1 billion by 2020 which is almost 70 percent of the world population. Since almost everyone, including passengers and drivers of private vehicles carry smartphone, our proposed solution takes the advantage of location based services that provides the devices geographic position in real time based on the information received from the satellites. The system uses the GPS(Global Positioning System)receiver, GSM(Global System for Mobile Communication) and internet connectivity services found in smartphones. Using GPS receiver, one can find the exact location, even set the navigation path from source to reach a particular destination. Internet connectivity (Wi-Fi/3G) plays a vital role to communicate with database residing in the remote server. By using the proposed system, catching a vehicle now just a click away without any cost of placing extra device other than smartphone.

The rest of the paper is organized as follows: Section II focuses on other existing related work on GPS service and smartphone, while Section III articulates about related technology. Section IV and V includes the architecture and flow of operation respectively to show the effectiveness of proposed system to rent private vehicles at real time. Implementation and evaluation are covered in section VI and VII correspondingly. Section VIII includes a discussion about the result. Concluded remarks and some future directions of our work are described in Section IX.

II. Related work

Several works have been done and going on GPS, GSM and internet connectivity features available in the smartphone. This section describes the various existing location based researches related to the proposed system, mainly, using tracking systems through GPS.

Le-Tien, T.; Vu Phung [2] describes a practical model for routing and tracking with mobile vehicle in a large area outdoor environment based on the Global Positioning System (GPS) and Global System for Mobile Communication (GSM). The system includes the Compass sensor-YAS529 of Yamaha Company and Accelerator sensor-KXSC72050 of Koinix Company to acquire moving direction of a vehicle. The system will acquire positions of the vehicle via GPS receiver and then sends the data to supervised center by the SMS (Short

Message Services) or GPRS (General Package Radio Service) service. The supervised center comprises of a development kit that supports GSM techniques-WMP100 of the Wavecom Company. Finally, the position of the mobile vehicle will be displayed on Google Map.

The paper presented by El-Medany, W.; Al-Omary et al [3] describes a real time tracking system that provides accurate localizations of the tracked vehicle with low cost. GM862 cellular quad band module is used for implementation. A monitoring server and a graphical user interface on a website is also developed using Microsoft SQL Server 2003 and ASP.net to view the proper location of a vehicle on a specific map. The paper also provides information regarding the vehicle status such as speed, mileage.

Sultana et. al. [4] proposed a smart, location based time and attendance tracking system, as manual attendance tracking system is time consuming and erroneous. It is a client server based android application where server is implemented on a personal computer using Apache- Tomcat 7. Here, the employees of the organization have to install the APK and save the office coordinates by entering the latitude, longitude, radius of area and IP(Internet Protocol) address of the office internet. When an employee enters or leaves his working area, the system sends the employee id and local time to the server using GPS and stores the information in a database.

Though the system reduces the cost of heterogeneous devices like electronic tags, bar code badges and magnetic stripe cards, it is developed for only android platform and if both the employee device and the whole system are not in the same internet connection(Wi-Fi/3G), it will provide erroneous result.

Shrestha R. et. al. [5] presented android-based location and message sharing system which provides a secured two way communication between web server and android based application using symmetric cryptography. Symmetric cryptography ensures the protection of data as existing android system has no centralized database and fails to protect the privacy of data causing the problem of data management and portability. The system uses Java programming language for android mobile application, PHP programming language to implement web server, MySQL as external database to store the data and JSON as intermediary between android platform and device.

G. B. Al-Suwaidi et. al. [6] proposed a client-server architecture based application Locating Friends and Family Using Mobile Phones with Global Positioning System (GPS) that helps the users to locate their family members and receive alerts when their friends are nearby. The mobile application was implemented using J2ME combining the most recent APIs and other older APIs together in order to make the application reliable on all types of mobiles. The server was implemented using PHP scripting language along with MySQL database.

Tekawade et. al. [7] created a mobile tracking application for locating friends using LBS(Location Based Services) to locate the user, track his friends and receive an alert message when nearby, based on radius set by administrator. By using GPS it helps the users to reach his destination without any suffering. To ensure the security of the user, it provides anti-theft facility so that his location information is sent to geographically the nearest police station. The system used J2ME for developing fronted application and database was maintained by MYSQL.

Turhan et. al. [8] developed a smartphone application to locate the nearest available blood donor volunteers and establish a communication with him/her, if the blood shares are insufficient. An uninterrupted connection is built between the donors and volunteers by updating the location information of available donors to main system and sending blood request to the donors. The application is written in JAVA for Android Operating system along with SQLite database to store persistent data using Android Studio IDE(Integrated Development Environment) and ANT build tool.

Here, we can see that many applications have been developed using GPS, GSM and internet connectivity features available in the smartphone. But a very few works in reality has been done about renting vehicles at real time to make our journey comfortable and safe. Our proposed system tries to complete this lacking.

III. Related Technology

To build the system described in the given sections, we have taken help of different technologies. In this stage we familiarize the readers with our system related technologies namely GPS, Android and LAMP to comprehend the essence of our paper.

A. GPS

24 satellites are systematized to form Global positioning system(GPS) and orbit the earth twice in a day maintaining a very precise perimeter. It provides the users positioning, navigation and timing facilities. Anyone can take these facilities without any subscription fee. GPS consists of mainly three segments:

1. The space segment
2. The control segment
3. The user segment

The space segment [9] consists of 24 satellites. Theses GPS satellites are powered primarily by sun-seeking solar panels, with nicad batteries providing secondary power. On board each GPS satellite are four atomic clocks, only one of which is in use at a time. These highly accurate atomic clocks enable GPS to provide the most accurate timing system that exists.

The control segment checks the position, speed of the satellites and transmits the updated navigational data. The satellites get these updated data in the signals and send to GPS receivers.

The user segment consists of the GPS receivers that detect, decode and process the signals received from the satellites and uses these informations to calculate users exact location.

B. Android

The Android operating system [10] is composed of a virtual machine that runs on the Linux kernel, plus APIs and a collection of built-in applications. It is open source and Google releases the code under the Apache License. This open source code and permissive licensing allows the software to be freely modified and distributed by device manufacturers. Additionally, Android has a large community of developers writing application written primarily in a customized version of the Java programming language [11].

C. LAMP

LAMP(Linux, Apache, MySQL and Perl/Python/PHP) is a combination of Apache web server, MySQL database engine, the Hypertext Preprocessor (PHP) and other popular scripting languages such as Perl and Python. It is a powerful and robust development environment for development and deployment of web based applications. Apache is the most commonly used web server on Linux system. Web servers are used to serve web pages using Hypertext Transfer Protocol(HTTP). PHP is a server side scripting language designed for web development but also used as a general purposed programming language. PHP code may be embedded into HTML code. The standard PHP interpreter, powered by the Zend Engine, is free software released under the PHP License. PHP has been widely ported and can be deployed on most web servers on almost every operating system and platform, free of charge [12].

D. Calculating distances using coordinates:

In order to calculate the distance between two points where the coordinates of each point is given, we have used haversine formula. The haversine formula is an equation important in navigation, giving great-circle distances between two points on a sphere from their longitudes and latitudes.[13]

For any two points on a sphere, the haversine of the central angle between them is given by:

$$hav(\frac{d}{r}) = hav(\phi_2 - \phi_1) + cos(\phi_1)cos(\phi_2)hav(\lambda_2 - \lambda_1)$$

where hav is the haversine function:

$$hav(\theta) = sin^2(\frac{\theta}{2}) = \frac{1 - cos(\theta)}{2}$$

Here,

- d is the distance between the two points
- r is the radius of the sphere
- θ_1, θ_2 are the latitude of point 1 and point 2 (in radians)
- λ_1, λ_2 are the longitude of point 1 and point 2 (in radians)

On the left side of the equals sign $\frac{d}{r}$ is the central angle, assuming angles are measured in radians. d is solvable by applying the inverse haversine (if available) or by using the arcsine (inverse sine) function:

$$d = rhav^{-1}(h) = 2r \arcsin(\sqrt{h})$$

where h is $hav(\frac{d}{r})$, or more explicitly d equals:

$$2r \arcsin(\sqrt{hav(\theta_2 - \theta_1) + cos(\theta_1)cos(\theta_2)hav(\lambda_2 - \lambda_1)})$$

We are considering it as equation number (1).

IV. Architecture of the Proposed System

The architecture of our proposed system, depicted in Fig. 1, indicates the main 5 elements that construct our system which are GPS, mobile clients, repository, server and map service.

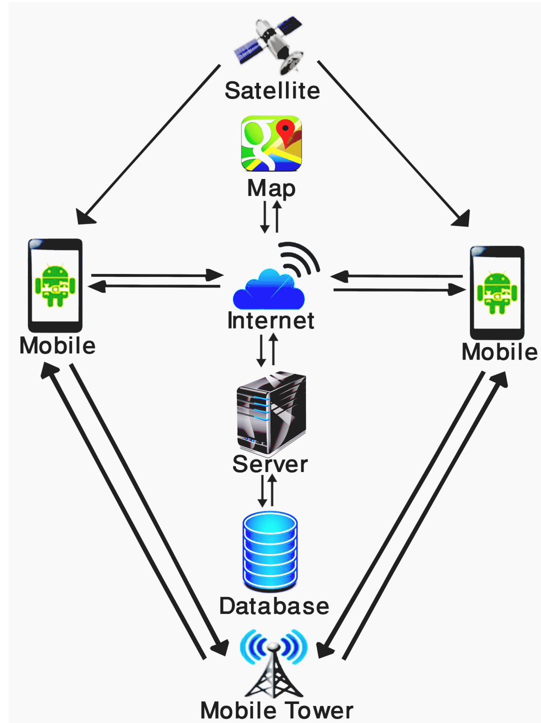


Fig. 1. Architecture of the proposed system

A. GPS satellites

The first and most important part of our system is tracking user's location and calculating the distance between the driver and the passenger to find the nearest vehicle. To fulfill these purposes, we have used GPS as it provides accurate user location. When the users change their location, the mobile client needs to update the user location in the server and it gets these information from GPS. The latitude and longitude of the user's location is determined by the GPS and in turn it sends them to the mobile client.

B. Mobile client

Mobile client means the mobile subscriber who is hooked up to the network by using particular service. Our mobile client consists of mainly two components :

1. GPS receiver 2. Android based smartphone

The inbuilt GPS receiver in smartphone, takes the signal information about the users current location from the GPS periodically and updates it in the server via internet.It uses trilateration method to calculate users exact location. After determining users exact location, it shows the list of the nearest drivers around the passenger and vice versa. When a passenger finds the nearest driver, he can make a voice call to the driver by using the GSM feature or give an SMS by using his phone. We have implemented the mobile client using Java Programming language on android operating system.

C. Repository

Repository is the aggregation of multiple databases to store various kinds of data for distribution over a network. The database of our system contains all the information about all subscribed users. It mainly consists of 4 attributes namely who, latitude, longitude and distance which stores users name, password, mobile number, latitude and longitude of users current location and updates the distance between the driver and the passenger by calculating these. We have chosen MYSQL to implement the database. Here we have used central database instead of having own database stored on the mobile phone to reduce the storage overhead on each phone.

D. Map service

Our system uses map service provided by Google. It makes an easy solution to view places before going there. It uses GPS to track the nearest vehicle around the user and provides street view between passenger and driver for finding proper spotPeople those who have come to a new city or are not familiar with every nook and corner of the road will find it useful.

E. Server

The architecture of our proposed system is server centric and the application server manages all information in the database. Passengers receive details information of all the drivers on their mobile from the server and vice versa. Internet works as the medium for transferring user data and service request from mobile to server and the requested information back to the user.

V. Flow of operation of the proposed system

In this section, we have described the overall methodology of our proposed system.

1. The application uses two user profiles, that of the passenger and of the driver. Both have to install the required system APK file into their Android device.

2. After starting the application, one will be asked whether registered user or not, as only the registered users can use this service. If not, he will be prompted to create account by providing username and password, phone number, mail address and whether s/he wants to use the system as a driver or passenger which will then be sent to the server to be saved in the central repository.

3. After successful registration, each time s/he wants to run the app,s/he has to log in with correct username and password as a proof of being authorised user. Then the system will check whether the mobile location service is on or not.If off, the system will prompt the user to turn it on.

4. Once the mobile location service is turned on, the system traces passenger's location, updates the location in the database through the server. Now, it calculates the distance between two mobile devices, that is, the driver and the passenger, to determine the nearest drivers. Then it shows a list of them on mobile clients.

5. Then the passenger send a rent request to a particular driver specifying the location where s/he wants to go to start a discussion. The driver receives the request and responses either affirmatively or negatively.

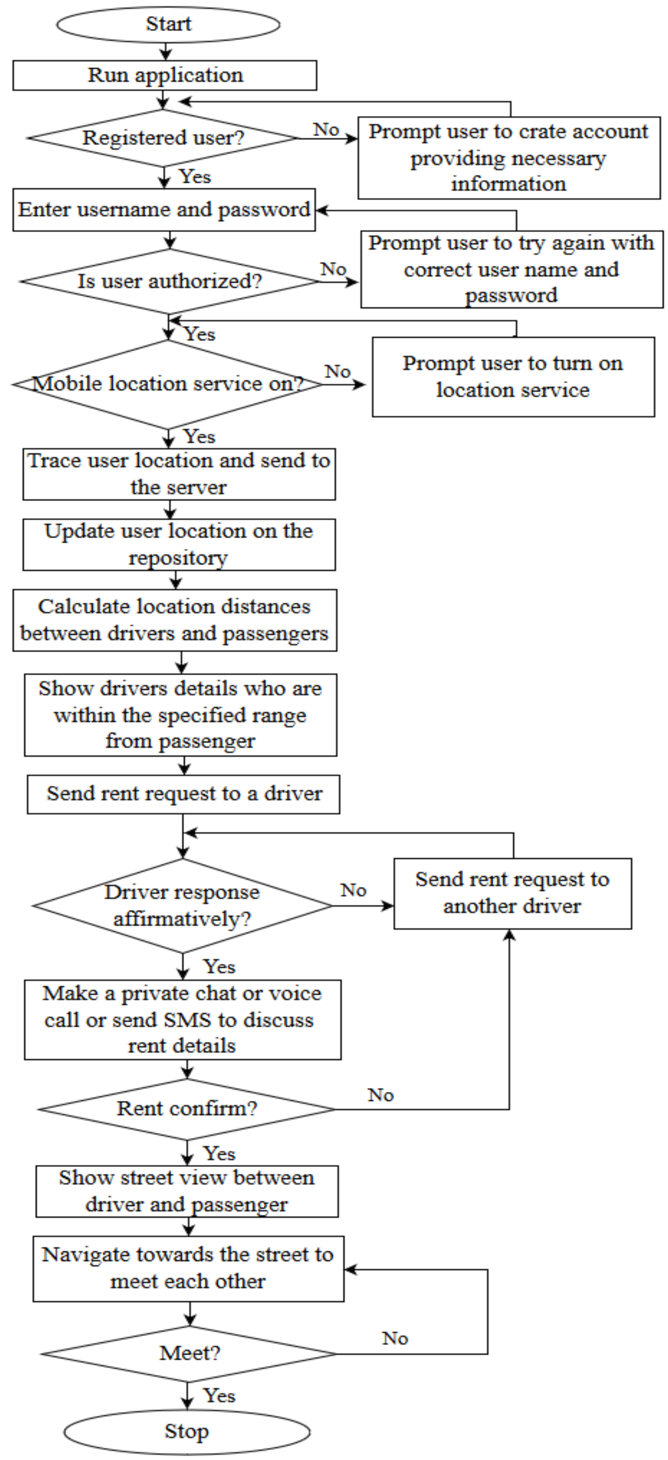


Fig. 2. Flow of operation of the proposed system

6. When the driver responses in the affirmative, the passengers makes a private chat or voice call or sends SMS to the driver about rent details. If the rent is not confirmed, the system allows to send request to another driver. After the rent

is confirmed, the street view is shown between the driver and passenger.

Figure 2 depicts the overall methodology of the proposed system.

VI. Implementation

As our project consists of a wide range of activities, all of them have not been fully implemented due to time constraint. The goal of strict implementation is nothing more than a proof that the concept generated here is completely achievable.

The mobile application was implemented using Java while PHP was used for implementing server functions and MYSQL was used for the database. It consists of five modules namely Registration/ Login module, GPS module, Conversation module, Showing street view module and Exit module. Due to time limitation, we have not properly implemented the 'Showing street view' module.

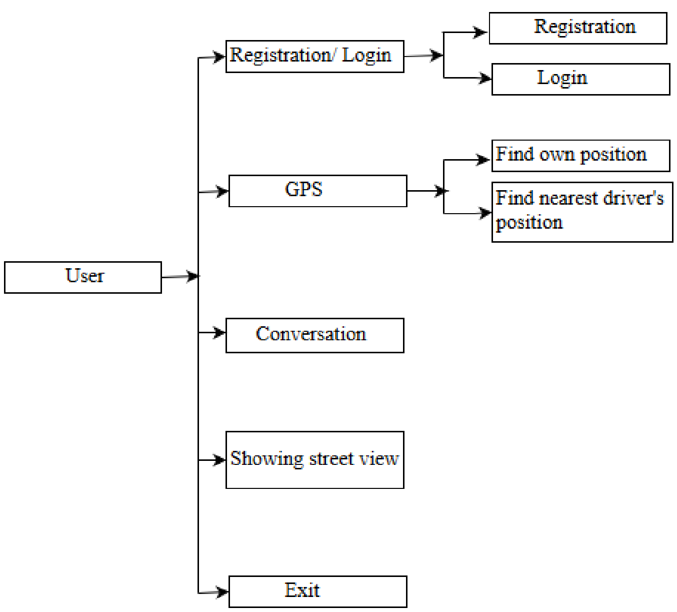


Fig. 3. Functional module of system

VII. Evaluation

This part gives an evidence about the average location accuracy of our application. Here, we have defined two points (named P and Q) on the ground and measured the distance between the two points using a measurement tool. Again, we used GPS receiver to get the coordinates of points and calculate the distance using Equation 1.

Points	Latitude	Longitude	Distance
Point P	21.34582	91.83303	20.246 m
Point Q	21.34565	91.83310	

Table 1: Distance calculation between two points (Point P & Point Q)

The measured distance using measurement tool was 22 meters while using GPS coordinates gave a distance of 20.246 meters, so the error in measuring the distance was 1.754 meters[6].

VIII. Result and Discussion

To check the user friendliness and practicability of this system, we have done a survey among 80 users to try this and received positive feedback from almost all of them. This survey was manipulated by asking several questions like "how quickly they have found vehicles by using this application? Does it provide accurate information about the location of the nearest vehicles? Do they think their personal information is at risk?" and so on. According to users, it is a handy device and also feasible. They have also found it as times saving and think that it will serve as a saviour when emergency appears.

But almost all of them are concerned with location privacy. Because without proper protection, the location information generated by our system can be abused in almost any domain of human, social or economic activity. But our system guarantees users' safety because all data are exchanged by taking users' consent. Without users' permission, their current location will not be updated and drivers can not be able to know the current location of the users'.

We are also trying to develop policy-based initiatives for privacy protection, like PDRM(Personal Digital Rights Management), P3P(Privacy Preferences Project) and GeoPriv[14].

We have also got couple of suggestions regarding inclusion of estimated rent service and more user friendly interface.

IX. Conclusion

This paper introduces a smart, location based private vehicle renting system that can make an end of all sufferings of people in renting vehicles only through a click. It will change the transport service radically by facilitating both the passengers and drivers. Passengers can easily locate their nearby private vehicles, make a chat on rent and rate a driver depending on his driving skill. Drivers can also find nearby passengers without loitering here and there.

The application gets the location coordinates of passengers plus drivers using GPS(Global Positioning System) as location provider and show that location on mobile client along with the street view between the passenger and the driver using Google map service.

The mobile client is implemented using Java Programming language on android operating system, server side programming is done using PHP and repository is maintained by MYSQL.

Finally, like any software product or design, there is still room for enhancement. Features can be added to enhance the system such as Geo-fencing, emergency alerts [15] and many others. Our future extensions can be summarized as follows:

1. Design a more user friendly graphical interface so that even a layman can use it without any confusion.

2. Provide an estimated rent by considering the distance between the passenger and his destination so that drivers can not take advantage of the ignorance of the passengers.

3. Propose newer algorithm to improve accuracy and lower power consumption.

4. Develop this system for windows and iOS platform.

REFERENCES

[1] ”6.1B Smartphone Users Globally by 2020, Overtaking Basic Fixed Phone Subscription” available at: http://techcrunch.com/2015/06/02/6-1b-smartphone-users-globally-by-2020-overtaking-basic-fixed-phone-subscriptions/ [Date accessed: March 7, 2016]

[2] Le-Tien, T.; Vu Phung-The, ”Routing and Tracking System for Mobile Vehicles in Large Area,” Electronic Design, Test and Application, 2010. DELTA ’10. Fifth IEEE International Symposium on , vol., no., pp.297,300, 13-15 Jan. 2010.

[3] El-Medany,W.;Al-Omary,A.;Al-Hakim,R.;Al-Irhayim,S.;Nusaif,M.,”A Cost Effective Real-Time Tracking System Prototype Using Integrated GPS/GPRS Module,” Wireless and Mobile Communications (ICWMC), 2010 6th International Conference on,vol.,no.,pp.521,525,20-25 Sept.2010

[4] Shermin Sultana, Asma Enayet, Ishrat Jahan Mouri, ”A smart, location based time and attendence tracking system using android application”, in International Journal of Computer Science, Engineering and Information Technology (IJCSEIT), vol. 5, no. 1, 2015.

[5] Ramesh Shrestha, Yao Aihong, ”Design of Secure Location and Message Sharing System for Android Platform”, in Computer Science and Automation Engineering (CSAE), 2012 IEEE International Conference on, vol. 1. IEEE, 2012, pp. 117121.

[6] G. B. Al-Suwaidi and M. J. Zemerly, ”Locating friends and family using mobile phones with global positioning system (gps)”, in Computer Systems and Applications, 2009. AICCSA 2009. IEEE/ACS International Conference on. IEEE, 2009, pp. 555558.

[7] A. Tekawade, A. Tutake, R. Shinde, P. Dhole, and M. S. Hirve, ”Mobile tracking application for locating friends using lbs”, in International Journal of Multimedia and Ubiquitous Engineering, vol. 7, no. 4, pp. 125140, 2013.

[8] S. TURHAN, ”An android application for volunteer blood donors”.

[9] ”Understanding the Global Positioning System (GPS)” available at: http://www.montana.edu/gps/understd.html [Date accessed: March 7, 2016]

[10] ”AT and T Developer Program” available at: http://developer.att.com/developer/forward.jsp?passedItemId=100036 [Date accessed: March 7, 2016]

[11] ”Android” available at: http://adityasaware.com/android.html [Date accessed: March 7, 2016]

[12] ”PHP-Wikipedia, the free encyclopedia” available at:http://en.wikipedia.org/wiki/PHP [Date accessed: March 7, 2016]

[13] ”Haversine formula-Wikipedia, the free encyclopedia” available at: https://en.wikipedia.org/wiki/Haversine formula [Date accessed: March 7, 2016]

[14] Matt Duckham, Lars Kulik, ”Location privacy and location-aware computing”

[15] Sharan Shetty, Nitin Ranjan, Aditya Sharma, Aman Prakash Singh, Rahul Koti, ”Surveying Smartphone Security for Linux Powered Devices”, in International Journal Of Engineering And Computer Science, ISSN: 2319-7242, Volume 4, Issue 11, Nov 2015, Page No. 15000-15004

Review on Telemonitoring of Maternal Health care Targeting Medical Cyber-Physical Systems

Mohammod Abul Kashem*‡, Md. Hanif Seddiqui†‡,
Nejib Moalla‡, Aicha Sekhari‡, Yacine Ouzrout‡

*Dept. of Computer Science and Engineering
Dhaka University of Engineering and Technology, Dhaka, Bangladesh

†Dept. of Computer Science and Engineering
University of Chittagong, Chittagong, Bangladesh

‡DISP Laboratory
IUT Lumiere Lyon 2, Lyon, France

Email: drkashemll@duet.ac.bd, hanif@cu.ac.bd,
nejib.moalla@univ-lyon2.fr, aicha.sekhari@univ-lyon2.fr, yacine.ouzrout@univ-lyon2.fr

***Abstract*—We aim to review available literature related to the telemonitoring of maternal health care for a comprehensive understanding of the roles of Medical Cyber-Physical-Systems (MCPS) as cutting edge technology in maternal risk factor management, and for understanding the possible research gap in the domain. In this regard, we search literature through google scholar for published studies that focus on maternal telemonitoring systems using sensors, Cyber-Physical-System (CPS) and information decision systems for addressing risk factors. We extract 1340 contents that address different maternal health issues. Among the large number of contents, we have elaborated 26 prospective articles relating to sensor and Medical Cyber-Physical-Systems (MCPS) based maternal telemonitoring. The short-listed articles are categorized as follows: 12 articles for maternal risk factors, 9 for synthesis matrices and 5 for other essential factors. From our review, we have extracted different vital symptoms as maternal risk factors during pregnancy. Moreover, we have identified a number of cyber-frameworks as the basis of information decision support system to cope with the different maternal complexities. We have observed the Medical Cyber-Physical System (MCPS) as a promising technology to manage the vital risk factors quickly and efficiently by the care provider from a distant place to reduce the fatal risks. Despite communication issues, MCPS is a key-enabling technology to cope with the advancement of telemonitoring paradigm in the maternal health care system.**

Keywords*—*Maternal Risk Factors, Medical Cyber-Physical-System, Critical System, Sensor network, Telemonitoring

I. Introduction

The fact sheets of World Health Organization (WHO) [1] contain the real scenario of the present maternal mortality status as a partial fulfillment of the Millennium Development Goals (MDG) 4 and 5. WHO announces a new agenda, called *Sustainable Development Goal (SDG)*, as a follow-up of MDG. The key target of the new agenda is to reduce the global maternal mortality ratio to less than 7 per 100,000 between 2016 and 2030. Now it is getting many researchers attention to contribute to the reduction of maternal mortality rate through introducing new technologies in the pathway.

With the recent advancement of Information and Communication Technology (ICT), especially wireless sensor networks (WSN), medical sensors and cloud computing, researchers pave the way for new attention to heterogeneous data integration and automation in the health care domain to attain the SDG, though isolated sensors such as ElectroCardioGram (ECG), UltraSonoGram (USG), echo etc. are being used in this domain since long ago. Sensors usually produce lots of data instantaneously with important and unimportant data as well. Automation often requires filtering unimportant data at the physical level to ease communication technology through transferring only important data to be used in the cyber (computer) level. This technique is coined as Cyber-Physical-Systems (CPS). CPS research is revealing numerous opportunities and challenges in medicine and biomedical engineering [1]. CPS in medicine and biomedical engineering is often called as *Medical Cyber-Physical Systems (MCPS)*.

Cyber-Physical-Systems (CPS) was identified as a key research area by the US National Science Foundation (NSF) in 2008 and was listed as the number one research priority by the US Presidents Council of Advisors on Science and Technology [2]. The advancement of WSN, medical sensors and cloud computing may enable CPS a powerful candidate for in-home patient care [3].

CPS in general health care domain has been reviewed [4] recently, however the study of CPS in maternal health care for telemonitoring is yet another important issue. As health care mission is critical services, we need to focus on the safety, security and sustainability (S3) issue in this domain [5]. Moreover, maternal health care systems are heavily affected by the risk factor identification and integration to produce an accurate alarm. Many issues are still open for future research

[1] http://www.who.int/mediacentre/factsheets/fs348/en/

such as real-time processing, efficient data query, storage management, and S3. Therefore, one of the aims of this review paper is to indicate some unanswered questions or research gap in CPS for health care.

In this paper, we have studied different systems, used technologies, various vital risk factors, and qualitative abilities. We have investigated different systems in terms of maternal and fetal monitoring, telemonitoring, storage, and Decision Making System (DMS), while technologies have been elaborated by means of CPS, sensors, communication. Various influencial maternity health status are orserved through external factors, environmental factors, vital risk factors, physical activity, and behavioral factors. Moreover, privacy, security, interoperability, adaptability, and reliability have been portrayed in this paper as essential qualitative ability of a system.

The rest of the paper is organized as follows. **Section II** adheres the literature search, study selection, data extraction and analysis. **Section III** contains maternal health-related detailed literature review that includes risk factors, synthesis matrix and other literature contribution whereas discussion on some interesting findings has been articulated in **Section IV**. Concluded remarks of our work are elaborated in **Section V**.

II. Materials and Methods

During conducting this review, we adhere to be focused on our targeted keywords: maternal health care, maternal risk factors, CPS or sensor.

A. Literature Search and Study Selection

We have conducted a comprehensive search through google scholar for extracting literature on CPS/sensor based telemonitoring system considering maternal risk factors from 2010 through 2015. During our search, we have used free text (e.g. CPS sensor maternal risk factors) to maximize the coverage. We have customized our search by putting range from 2010 through 2015 that includes patent and citation in google scholar. We have found 1340 articles from the search in the google scholar [2].

Fig. 1 summarizes our process of literature search and study selection. Initially, we have found 1340 articles through google scholar with the free query text: CPS sensor maternal risk factors. At the primary scrutinizing, we have eliminated duplicate, non-relevant, and non-technical articles. We have detected 173 relevant articles and have focused these articles for further deeper analysis. In this stage, 60 review or editorial articles, reports, case studies and a few not really relevant articles have been eliminated. Then we have remained 113 articles. We have gone through the abstract of each article to detect interesting findings and have found 87 papers non-relevant due to the lack of CPS/sensor, absent of using Information and Communication Technology (ICT) or not addressin maternal risk factors. Finally, we have 26 prospective studies that focus on our concentrated field of interest.

[2]https://scholar.google.fr

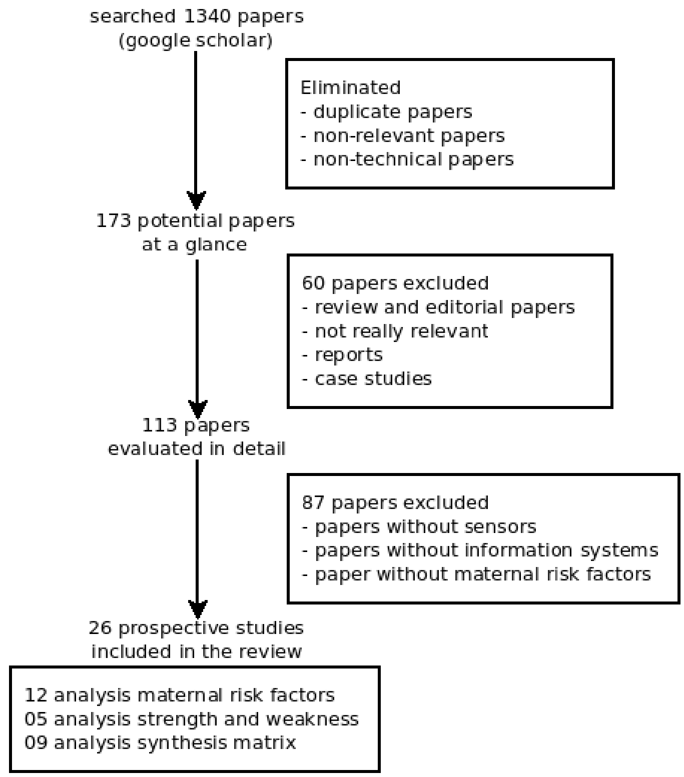


Fig. 1. Flow diagram for literature search and study selection

B. Data Extraction and Analysis

The final 26 articles are classified into three groups: 1) 12 papers are studied to identify vital risk factors associated to maternal health care, 2) 5 articles are analyzed to focus on other essential factors to pinpoint the research gap, and 3) 9 articles are analyzed to find attributes that can characterize any other literature.

We have assessed 12 articles to identify risk factors and have found the following risk factors evidently: age, body mass index(BMI), blood pressure (BP), blood oxygenation (BO), blood glucose (BG), body temperature, physical activity, maternal ECG (MECG), nausea at the first-trimester, vaginal discharge at the first-trimester, maternal serum alphafetoprotein level (MSAFP) at the second-trimester, contraction at the third-trimester, abnormal fetal position, electrical uterine activity (EUA), mechanical uterine activity (MUA), fetal heart rate (FHR), and fetal movement activity.

In the second group of 5 articles, we have noticed authors' claims and considered as their strength to find some interesting key-enabling ideas.

We have detected a number of attributes to characterize any relevant articles. The attributes are as follows: CPS, sensor technology, telemonitoring, interoperability, S3 (safety, security and sustainability), adaptability, reliability, cyber repository or storage, external or environmental factors, communication technology, DMS, risk factors, physical activity, behavioral factors, privacy issues, maternal monitoring and fetal monitoring.

III. Maternal Health Review Result

We have analyzed the articles in three aspects: risk factors, synthesized matrices, and other contributions.

A. Risk Factors

As we concentrate our study on maternal health care, we look for fundamental causes or risk factors of maternal health care. The risk factors include age, body mass index(BMI), blood pressure (BP), blood oxygenation (BO), blood glucose (BG), body temperature, physical activity, maternal ECG (MECG), nausea at the first-trimester, vaginal discharge at the first-trimester, maternal serum alpha-fetoprotein level (MSAFP) at the second-trimester, contraction at the third-trimester, abnormal fetal position, electrical uterine activity (EUA), mechanical uterine activity (MUA), fetal heart rate (FHR), fetal movement activity and so on. All the factors have threshold values or measurement units. Medical professionals or researchers research to find out the different factors for further medication, clinical diagnosis or other management. Due to the different factors, a pregnancy can be classified as normal, moderate or high-risk pregnancy by an obstetrician [6].

In the paper [7], authors study on the range of maternal ages and find out the lower and upper bound of age for high-risk pregnancy. They have shown that the advanced maternal age is associated with a range of adverse pregnancy outcomes. Moreover, a study shows a method for comparing age factor of a pregnant woman and a BMI model to classify as underweight ($< 18.5kg/m^2$), normal ($18.524.9kg/m^2$), overweight ($25 - 29.9kg/m^2$), obese ($30 - 34.9kg/m^2$) and morbidly obese($> 35kg/m^2$) [8].

A report [9] and a study [10] classify hypertensive disorders of pregnancy into following categories: gestational hypertension, chronic hypertension, pre-pre-eclampsia, and pre-eclampsia superimposed on pre-existing hypertension. According to the international guidelines, the treatment of hypertension in pregnancy varies with respect to thresholds.

Authors in the paper [11] develop a method to fix up SpO2, *defined as Oxygen saturation by pulse Oximetry*, value for pregnant woman with pre-eclampsia by PIERS (Pre-eclampsia Integrated Estimate of RiSk). They discover threshold values of SpO2 as $\leq 93\%$ which confers the particular risk.

Diabetes in pregnancy is well studied in the paper [12], which proposes new diagnostic criteria for gestational diabetes and finds fasting plasma glucose:5.1 mmol/l, 1 h plasma glucose:10.0 mmol/l, 2 h plasma glucose:8.5 mmol/l. The authors also discuss different management procedure of diabetes and mention that diabetes may not be harmful to a pregnant woman if it can be monitor properly, otherwise, gestational diabetes may become a big issue comparatively with other risk factors of a pregnant woman.

Hyperthermia has been studied in [13] to demonstrate an association between high maternal fever in early pregnancy and Neural Tube Defects (NTDs). The author makes experiments with hot tub or spa used by pregnant women and find a threshold body temperature value and maintained below 38.9C.

In the paper [14], authors survey through physical activity questionnaires and find an interesting correlation of BMI and Gestational Weight Gain (GWG). Women with higher BMI have a larger decline in physical activity.

A researcher develops Maternal Serum Screening (MSS) methods by using different case study. He has suggested MSS level for Distribution of MSAFP levels with fetal neural tube defects as 7.0 (unaffected), 2.3(Spina Bifida), 5.0(Anencephaly) and MSS level for distribution of MSAFP levels with fetal Down syndrome as 0.5/LR-2.0(Down syndrome), 0.8/LR-1.0,1.4/LR 0.5(unaffected) during the second trimester of pregnancy period [15].

In the paper [16], authors develop a new method to detect the uterine contraction by using changes of several ElectroMyoGraphy (EMG) parameters. They have considered the parameters power spectrum (PS) peak frequency and propagation velocity(PV) to find out a summation value as a cut-off. They observe an interesting finding that if the value exceeds the cut-off as 84.48, the delivery will be within 7 days. They have experimented against 100 women and theoretically got a true labor time.

Fetal heart rate is observed in [17] with a methodology based on the “delayed moving windows” algorithm. The researchers have determined the normal fetal heart rate as 120 to 160 bpm.

B. Synthesis Matrices

Fig. 1 summarizes our process of literature search and study selection based inclusion and exclusion criteria to select 9 prospective studies for detailed synthesis matrices. We have selected the 9 contributions as they define a full system to address maternal health care issues. The synthesis matrices are developed based on the meta-information of the articles, different systems used in their proposal, technology usage, different maternal factors and qualitative attributes.

We summarize the meta information of the contributed articles in Table I, which depicts the fact that the Cyber-Physical-System based maternal health care systems have been widely studied in USA, European countries and China.

TABLE I. Meta-information on Selected Publications

Source	Publication Year	Authors Location
mMonitoring [18]	2013	China
homeCPS [19]	2014	Poland
wSensors [20]	2015	USA, The Netherlands, Belgium
fMDU [21]	2014	Poland
mMamee [22]	2015	Greece, UK
bsAcquisition [23]	2015	Poland
health-CPS [24]	2015	Saudi Arabia, China, USA, Taiwan
hTelecare [25]	2015	Poland
iWSN[26]	2015	Romania

In Table II, we have observed different studies with maternal and fetal monitoring, telemonitoring, used storage and Decision Making System (DMS) attributes to elaborate properties of each of the articles. The table depicts that all of the reviewed systems have maternal telemonitoring systems with Maternal Heart Rate (MHR), Electrocardiogram (ECG) and other common monitoring factors. However, two articles do not take fetal monitoring into account. Fetal monitoring often observes Fetal Heart Rate (FHR), ECG, Cardiotocography (CTG), QRS (three graphical deflections on ECG), fetal movement, Generalized Singular Value Decomposition (GSVD), ST segment ANalysis (STAN), and so on. Moreover, each study focuses on different monitoring factors along with different DMS. Interestingly, almost half of the studies use cloud storage, whereas the rest half of the studies use internet enabled hospital system.

We have attributed technology with CPS, sensors and communication technologies. Our observation on the studies is articulated in Table III. Almost half of the studies have already started using Medical Cyber-Physical-Systems, while all of them are using different sort of sensors such as Body Area Network (BAN), Personal Area Network (PAN), Micro-Detectors (MD), Maternal ECG (MECG), Fetal ECG (FECG), electrohysterography (EHG) and so on. Interestingly, bluetooth is quite common as a sensor communication technology.

There are a number of factors that can affect maternal health. These factors are characterized as external, environmental, risk, physical and behavioral factors. External factors often address eating habits, maternal activities, and the geographic location at which a pregnant woman is staying instantaneously while environmental factors are temperature, humidity, noise, the amount of CO_2 in the atmosphere, water, dust and etc. Risk factors are the main focusing points for maternal pregnancy risk calculation. However, behavioral factors such as sleeping pattern, stress, diet and weight management, smoking and drinking have hazardous relation with fetal development and maternal health. Our observation on different studies compares the facts in Table IV.

At last but not least the synthesis matrix focuses on the qualitative analysis of our reviewed articles. Among the qualitative factors, privacy stays on the top to make such personal health care system successful. As the health care is a critical issue, it needs a high level of safety, security and reliability. Moreover, as the maternal health care has a dynamic behavior over time, telemonitoring often requires self-adaptability to make this a success. After all, maternal health care often requires many sensors, technologies and systems to work together. Therefore, interoperability plays an important role in this critical issue. The comparison has been depicted in Table V to point on some interesting research gaps.

C. Other contributions

Our study also observes the strength of a few other articles in the discourse such that the strength can be addressed during designing a new system.

Authors in the article [5] articulate S3, *defined as safety, security and sustainability*, where safety focuses on the avoidance of hazards, security on the assurance of integrity, authenticity, and confidentiality of information, while sustainability addresses the maintenance of a long-term operation of CPSs using green sources of energy.

The article [2] introduces CPeSC3 (Cyber Physical enhanced Secured wireless sensor networks in 3 Cores) using integrated cloud computing in a medical health care application scenario focusing on 3C defined as, 1.communication core, 2.computation core and 3. resource scheduling and management core.

Privacy issue in maternal health care system has a prevalent impact on societal acceptance. This issue has been widely studied in [33]. This study focuses on the services for pregnant women at anytime and anywhere, with anything and anyone, where the actors are pregnant women, gynecologists, and caregivers.

The article [34] focuses on a *real-time* intelligent system that uses data mining technique against continuous sensor data kept in data center to produce service delivery. This study takes distributed file system and privacy issue into account.

In the article [6], authors address a machine learning approach for the early detection of the pregnancy risk based on patterns gleaned from a profile of known clinical parameters. It can predict and detect early complications of pregnancy using clinical decision support system (CDSS), which uses classification and regression trees to evaluate the risk category of a pregnancy.

IV. Discussion

As we have concentrated our study on maternal health care, we observe the fundamental risk factors of maternal health care. An author group, K. Horoba et al have done a number of research on the maternal health care considering Medical CPS and addressing important risk factors related to a pregnant woman. To develop an MCPS based monitoring system for maternal health care, it is always necessary to consider the system as a critical application. Every pregnant woman may have any two, more risk complexities, which are usually combined to produce a final alert to the patient as well as to the care -providers. As a critical domain, it often needs an intervention of medical professionals.

Through our analytical review over risk factors, we have found a research paper [6] that focuses on the automatic risk assessment tool for pregnancy care using Clinical Decision Support System (CDSS). They have addressed a process of combining pregnancy risk factors to classify synthetically the instantaneous pregnancy states as normal, moderate or high-risk pregnancies. Our study also focuses on all of the respective risk factors to find an interesting correlation.

Patients of pregnancy are called to be at high-risk if she or her baby has an increased chance of a health problem. There are thousands of risk factors for pregnant woman categorized as physical, behavioral, environmental, lifestyle factors. As our goal of this work is to study on maternal health care targeting Cyber-Physical-Systems, we analyzed especially the risk factors related to the pregnancy.

The recent advancement of medical devices, intelligent sensors, Internet of Things (IoT), efficient telecommunication and information based smart decision support system (DSS) has paved a paradigm shift in maternal health care. However, it necessitates the data filtering at a physical level, automatic integration of filtered information before taking a decision, avoiding data hazards, assurance of integrity, authenticity, and confidentiality of information, data failure prevention, sustainability using green technology are the key ideas to make a great success of the critical mission in maternity health care telemonitoring systems targeting Medical Cyber-Physical-Systems.

TABLE II. STATE OF THE ART SYSTEMS

Source	Maternal Monitoring	Fetal Monitoring	Telemonitoring	Storage	DMS
mMonitoring [18]	continuous	ECG, CTG, STAN	yes	cloud	yes
homeCPS [19]	on demand	FHR, QRS, movement, ECG	yes	hospital surveillance center	yes
wSensors [20]	yes	-	yes	-	machine learning and pattern recognition
fMDU [21]	cardiotocograph	FHR, amniotic fluid	-	-	Butterworth filter
mMamee [22]	yes	-	yes	cloud	analysis and correlation
bsAcquisition [23]	GSVD	FHR, movement CTG,QRS	yes	hospital surveillance center	ICA, abdominal signal analysis
health-CPS [24]	MHR and electrocardiogram	-	individual vital sign	cloud	yes
hTelecare [25]	Uterine activity, blood glucose MHR, blood pressure and oxygenation	FHR, ECG	yes	hospital surveillance center	automated quantitative analysis
iWSN[26]	uterine contractions,respiration rate Pulse oximetry, hemoglobin in blood	FHR	yes	cloud	wavelet signal analysis

TABLE III. USED TECHNOLOGY IN DIFFERENT SOURCES

Source	CPS	Sensors	Communication
mMonitoring [18]	-	Wearable Wireless	serial port, bluetooth, Zigbee, RF433, smart-phone
homeCPS [19]	MCPS	BAN, PAN, maternal and fetal, external and environmental	Bluetooth, Zigbee, GSM, Internet PDA, Laptop
wSensors [20]	-	Mobile and wearable sensors	-
fMDU [21]	-	Doppler ultrasound	-
mMamee [22]	-	Smart sensors	ZigBee, bluetooth
bsAcquisition [23]	MCPS	BAN,PAN, biosignal networked MD, MECG, FECG, EHG	ZigBee, bluetooth
health-CPS [24]	MCPS	Wearable BAN	Internet, bluetooth
hTelecare [25]	MCPS	BAN, PAN, MD	WAN/Internet
iWSN[26]	-	Mobile cardiotocograph and body sensors, Lilypad	Internet, wireless network, bluetooth, smart-phone

V. CONCLUSION

In summary of this review paper, our observations were of three folds. We observed different risk factors associated to maternal health care during pregnancy, while we studied comparisons of a few proposed systems considering meta-information of the articles, system, technology, factors and qualitative attributes. Moreover, we observed a few other well-designed prospective studies with the longitudinal assessment to find out some more interesting attributes to be focused on Medical Cyber-Physical-Systems (CPS) based maternal health care telemonitoring system.

ACKNOWLEDGEMENT

This work was supported by the European Erasmus Mundus cLINK programme at the University Lumiere Lyon 2, France.

REFERENCES

[1] R. Baheti and H. Gill, “Cyber-physical systems,” *The impact of control technology*, vol. 12, pp. 161–166, 2011.

[2] J. Wang, H. Abid, S. Lee, L. Shu, and F. Xia, “A secured health care application architecture for cyber-physical systems,” *Journal of Control Engineering and Applied Informatics*, vol. 13, no. 3, pp. 101–108, 2011.

[3] A. Milenković, C. Otto, and E. Jovanov, “Wireless sensor networks for personal health monitoring: Issues and an implementation,” *Computer communications*, vol. 29, no. 13, pp. 2521–2533, 2006.

[4] S. Haque, S. Aziz, and M. Rahman, “Review of cyber-physical system in healthcare,” *International Journal of Distributed Sensor Networks*, vol. 2014, 2014.

[5] A. Banerjee, K. Venkatasubramanian, T. Mukherjee, and S. Gupta, “Ensuring safety, security, and sustainability of mission-critical cyber–physical systems,” *Proceedings of the IEEE*, vol. 100, no. 1, pp. 283–299, 2012.

[6] A. Gorthi, C. Firtion, and J. Vepa, “Automated risk assessment tool for pregnancy care,” in *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE.* IEEE, 2009, pp. 6222–6225.

TABLE IV. MATERNAL DIFFERENT INFLUENCIAL FACTORS

Source	External Factors	Environmental Factors	Risk Factors	Physical activity	Behavioral Factors
mMonitoring [18]	-	-	ECG, BG, BP, oximeter, weight, and fat monitors	-	-
homeCPS [19]	maternal activity	temperature, humidity	BP, oxygenation,BG, body temperature, activity, MECG, EUA,MUA, FHR, FECG, fetal movement	yes yes yes	-
wSensors [20]	yes	-	obesity, BP, BG, stress, anxiety, sleep disorder	yes	sleep [27], stress [28], diet and weight management [29], [30] smoking [31], drinking [32]
fMDU [21]	-	-	fetal movement activity	-	-
mMamee [22]	yes	Temperature, humidity, CO_2, noise, water, dust	external factors	-	-
bsAcquisition [23]	maternal activity	temperature, humidity	diabetes, post-term pregnancy, pregnancy induced hypertension	-	-
health-CPS [24]	eating habits	temperature, humidity	obesity and high blood pressure	yes	emotion
hTelecare [25]	maternal activity	temperature, humidity	diabetes, pregnancy induced hyper tension, post-term pregnancy, BMI	yes	behavioral model
iWSN[26]	-	-	hypoxia	yes	-

TABLE V. QUALITATIVE ANALYSIS OF DIFFERENT SOURCES

source	Privacy	Security	Interoperability	Adaptability	Reliability
mMonitoring [18]	-	-	yes	-	yes
homeCPS [19]	yes	-	yes	-	yes
wSensors [20]	yes	-	-	yes	yes
fMDU [21]	-	-	-	-	-
mMamee [22]	yes	-	yes	yes	yes
bsAcquisition [23]	-	-	yes	yes	yes
health-CPS [24]	yes	security tag	-	yes	yes
hTelecare [25]	-	yes	yes	yes	yes
iWSN[26]	-	-	-	-	-

[7] R. Lampinen, K. Vehviläinen-Julkunen, and P. Kankkunen, “A review of pregnancy in women over 35 years of age,” *The open nursing journal*, vol. 3, no. 1, 2009.

[8] L. Kenny, T. Lavender, R. McNamee, S. ONeill, T. Mills, and A. Khashan, “Advanced maternal age and adverse pregnancy outcome: evidence from a large contemporary cohort,” *PloS ONE*, vol. 8, no. 2, pp. e56 583–e56 583, 2013.

[9] N. H. B. P. E. Program, “Report of the national high blood pressure education program working group on high blood pressure in pregnancy,” *American journal of obstetrics and gynecology*, vol. 183, no. 1, pp. s1–s22, 2000.

[10] R. Mustafa, S. Ahmed, A. Gupta, and R. Venuto, “A comprehensive review of hypertension in pregnancy.” *Journal of Pregnancy*, vol. 2012, pp. 105 918–105 918, 2012.

[11] M. Douglas, J. Hutcheon, T. Lee, L. Magee, and K. Walley, “Oxygen saturation as a predictor of adverse maternal outcomes in women with preeclampsia,” *J Obstet Gynaecol Can*, vol. 33, no. 7, pp. 705–714, 2011.

[12] S. Ali and A. Dornhorst, “Diabetes in pregnancy: health risks and management,” *Postgraduate medical journal*, vol. 87, no. 1028, pp. 417–427, 2011.

[13] C. Chambers, “Risks of hyperthermia associated with hot tub or spa use by pregnant women,” *Birth Defects Research Part A: Clinical and Molecular Teratology*, vol. 76, no. 8, pp. 569–573, 2006.

[14] Z. Sui, L. J. Moran, and J. Dodd, “Physical activity levels during pregnancy and gestational weight gain among women who are overweight or obese,” *Health Promotion Journal of Australia*, vol. 24, no. 3, pp. 206–213, 2014.

[15] B. Chodirker, “Maternal serum screening: What do the results mean?” *Canadian Journal of Diagnosis*, vol. 18, no. 7, pp. 62–66, 2001.

[16] M. Lucovnik, R. Kuon, L. Chambliss, W. Maner, S. SHI, L. Shi, J. Balducci, and R. Garfield, “Use of uterine electromyography to diagnose term and preterm labor,” *Acta obstetricia et gynecologica Scandinavica*, vol. 90, no. 2, pp. 150–157, 2011.

[17] S. von Steinburg, A. Boulesteix, C. Lederer, S. Grunow, S. Schiermeier, W. Hatzmann, K. Schneider, and M. Daumer, “What is the normal fetal heart rate?” *PeerJ*, vol. 1, 2013.

[18] P. Lyu, M. Peng, Y. Lyu, Y. Chen, and J. Yang, “A multi-communication-fusion based mobile monitoring system for maternal and fetal information,” in *e-Health Networking, Applications & Services (Healthcom), 2013 IEEE 15th International Conference on.* IEEE, 2013, pp. 559–563.

[19] K. Horoba, J. Jeżewski, J. Wróbel, A. Pawlak, R. Czabański, P. Porwik, and P. Penkala, “Design challenges for home telemonitoring of pregnancy as a medical cyber-physical system,” *Journal of Medical Informatics & Technologies*, vol. 23, 2014.

[20] J. Penders, M. Altini, C. Van Hoof, and E. Dy, “Wearable sensors for healthier pregnancies,” *Proceedings of the IEEE*, vol. 103, no. 2, pp. 179–191, 2015.

[21] J. Wróbel, T. Kupka, K. Horoba, A. Matonia, D. Roj, and J. Jeżewski, “Automated detection of fetal movements in doppler ultrasound signals versus maternal perception,” *Journal of Medical Informatics & Technologies*, vol. 23, pp. 43–50, 2014.

[22] K. Karagiannaki, S. Chonianakis, E. Patelarou, A. Panousopoulou, and M. Papadopouli, “mmamee: A mhealth platform for monitoring and assessing maternal environmental exposure,” in *Computer-Based Medical Systems (CBMS), 2015 IEEE 28th International Symposium on.* IEEE, 2015, pp. 163–168.

[23] A. Pawlak, K. Horoba, J. Jezewski, J. Wrobel, and A. Matonia, “Telemonitoring of pregnant women at homebiosignals acquisition and

measurement," in *Mixed Design of Integrated Circuits & Systems (MIXDES), 2015 22nd International Conference*. IEEE, 2015, pp. 83–87.

[24] Y. Zhang, M. Qiu, C. Tsai, M. Hassan, and A. Alamri, "Health-cps: Healthcare cyber-physical system assisted by cloud and big data," *Journal of Medical Informatics & Technologies*, 2015.

[25] J. Wrobel, J. Jezewski, K. Horoba, A. Pawlak, R. Czabanski, M. Jezewski, and P. Porwik, "Medical cyber-physical system for home telecare of high-risk pregnancy: Design challenges and requirements," *Journal of Medical Imaging and Health Informatics*, vol. 5, no. 6, pp. 1295–1301, 2015.

[26] A. Niţulescu, M. Crişan-Vida, L. Stoicu-Tivadar, and E. Bernad, "Integrated wireless sensor network for monitoring pregnant women." *Studies in health technology and informatics*, vol. 210, pp. 354–358, 2015.

[27] G. Pien and R. Schwab, "Sleep disorders during pregnancy," *SLEEP-NEW YORK THEN WESTCHESTER-*, vol. 27, pp. 1405–1417, 2004.

[28] A. Huizink, P. Robles de Medina, E. Mulder, G. Visser, and J. Buitelaar, "Stress during pregnancy is associated with developmental outcome in infancy," *Journal of Child Psychology and Psychiatry*, vol. 44, no. 6, pp. 810–818, 2003.

[29] S. Thangaratinam, E. Rogozinska, K. Jolly, S. Glinkowski, W. Duda, E. Borowiack, T. Roseboom, J. Tomlinson, J. Walczak, R. Kunz *et al.*, "Interventions to reduce or prevent obesity in pregnant women: a systematic review," *Health Technology Assessment*, vol. 16, no. 31, pp. 1–191, 2012.

[30] J. Louie, T. Markovic, N. Perera, D. Foote, P. Petocz, G. Ross, and J. Brand-Miller, "A randomized controlled trial investigating the effects of a low–glycemic index diet on pregnancy outcomes in gestational diabetes mellitus," *Diabetes care*, vol. 34, no. 11, pp. 2341–2346, 2011.

[31] M. Sexton and J. Hebel, "A clinical trial of change in maternal smoking and its effect on birth weight," *Jama*, vol. 251, no. 7, pp. 911–915, 1984.

[32] G. Carson, L. Cox, J. Crane, P. Croteau, L. Graves, S. Kluka, G. Koren, M. Martel, D. Midmer, and I. Nulman, "Alcohol use and pregnancy consensus clinical guidelines." *Journal of obstetrics and gynaecology Canada: JOGC= Journal d'obstetrique et gynecologie du Canada: JOGC*, vol. 32, no. 8 Suppl 3, pp. S1–31, 2010.

[33] M. Talpur, M. Shaikh, R. Buriro, H. Talpur, F. Talpur, H. Shafi, and M. Shaikh, "Internet of things as intimating for pregnant womens healthcare: An impending privacy issues," *TELKOMNIKA Indonesian Journal of Electrical Engineering*, vol. 12, no. 6, pp. 4337–4344, 2014.

[34] L. Chen, J. Yang, Q. Wang, and M. Wang, "An intelligent information system for maternal and child health care," in *Computer Software and Applications Conference Workshops (COMPSACW), 2011 IEEE 35th Annual*. IEEE, 2011, pp. 441–445.

Recursive Suffix Stripping to Augment Bangla Stemmer

Md. Hanif Seddiqui*, Abdullah Al Mohammad Maruf*, and Abu Nowshed Chy†
* Department of Computer Science and Engineering
University of Chittagong
Chittagong-4331, Bangladesh
†Department of Computer Science & Engineering
Toyohashi University of Technology
Toyohashi, Aichi, Japan
hanif@cu.ac.bd, maruf.cu.cse09@gmail.com, and nowshed@kde.cs.tut.ac.jp

***Abstract*—Stemming is an operation on a word that simply extract the main part possibly close to the relative root, *we define as a lexical entry rather than an exact morpheme*, by eliminating the suffixes, as they mean, without doing a morphological analysis. Stemming is a widely used, however, a basic tool for a language processing in the area of Information Access (IA). Previous research on Bangla stemming has shown the capability of eliminating a single suffix on a 'longest match' basis from a word. Our proposed system has augmented the research through introducing a recursive process to eliminate multiple suffixes from a single word to retrieve more relevant relative root. Our proposed algorithm contains three approaches in recursive suffix stripping based stemming. We define them as *conservative*, *aggressive*, and *rule-based* approaches. In our proposed method, an inflectional word is stemmed in all possible ways by the recursive suffix stripping algorithm before identifying the final stem using the conservative, the aggressive and the rule-based approaches. Our experiments and evaluation on a relatively larger text corpora show the strength and efficiency of our proposed algorithm with 92% accuracy.**

***Keywords*—Bangla Stemmer, Natural Language Processing (NLP), Information Access, Information Processing**

I. Introduction

Natural language often contains inflectional terms or variants coming off a single lexical element, called stem. A stem is often associated with an article, number, and other suffixes to form morphological variances in a language. Stemming is an operation that splits a terms into its constituent main part and a single or a number of suffixes. Splitted main part is not necessarily be the lexical root rather than it is another lexical word close to the root. We call this main part of a word as a relative root. Therefore, stemming is a different technique than morphological analysis that splits a term into its 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 in information retrieval (IR) domain. It improves the performance through augmented term-weighting, document-weighting, query-document matching, document indexing, and retrieval with query terms. Bangla is a highly inflectional language that contains tens of morphological variances on a single stem. However, a few study has addressed this issue in Bangla Information Retrieval.

There are a number of stemming techniques to retrieve word-stem in English. One of the simplest technique utilize a list of frequent suffixes to reduce their relative roots [1], [2]. A more detailed evaluation of stemming algorithms [3] has revealed the fact of significant improvement in the recall due to stemming techniques against English documents. Other languages have similar stemmer to assist their IR area of research. For the Dutch language, the suffix stripping is evaluated to find out the effectiveness of stemming [4]. A detailed survey has been conducted on existing techniques of stemming Indonesian words to their morphological roots [5]. Moreover, Punjabi language stemmer [6], Malaysian Malay language stemmer [7] has been focusing on their languages. A more detailed survey on stemming of Asian languages has been addressed in [8]. Ramanathan and Rao (2003) propose a lightweight stemmer for Hindi which has performed longest match stripping with a handcrafted suffix list [9]. Other stemming algorithms for Information Retrieval (IR) applications have been developed for different languages including Latin [10], Indonesian [11], Swedish [12], Dutch, German and Italian [13], French [14], Slovene [15], Turkish [16] and so on.

The Bangla stemming algorithms strip the suffixes using a predetermined suffix list. Dasgupta and Ng [17] addresses unsupervised morphological parsing to segment words into prefix, suffix, and stem. Islam et al. [18] focus on a 'longest match' basis, whereas Majumder et al. [19] uses clustering approach to strip suffixes. Bangla is such an inflectional language that has a tendency of containing more than one suffices in a single stem. Therefore 'longest match' may be inadequate to eliminate all suffixes from morphological variances. To overcome this limitation, we propose a recursive process to extract all possible stems. In our proposed stemming technique, a candidate word is stemmed using the minimal list of non-redundant suffixes recursively, which produces a list of all possible stems for further analysis to detect a proper stem.

The rest of the paper is organized as follows. **Section II** compares our idea with other existing related work to articulate a research gap. **Section III** describes our proposed algorithm along with some comprehensive examples of Bangla stemming. **Section IV** includes experiments and evaluation to show the effectiveness of our proposed stemming technique against a relatively large corpus. Concluded remarks and some future directions of our work is elaborated in **Section V**.

II. Related Work

We have a few stemmer in Bangla: Dasgupta and Ng's morphological parser [17], a light weight stemmer of Islam et al [18]. and Yet Another Suffix Stripper (YASS) of Majumder et al [19].

Dasgupta and Ng (2006) proposed unsupervised morphological analysis parsed the root from Bangla words. It tried to segmented words into prefixes, suffixes and stems without language-specific prior knowledge of morpho-tactics and morpho-phonological rules. Authors defined the parser in two steps: inducing prefixes, suffixes and roots from a vocabulary of a large, unannotated corpus, and segmenting morphological variances based on these induced morphemes. The authors evaluated their system against 4,110 human-segmented Bangla words, and they claimed that their algorithm obtained F-scores of 83% that substantially outperformed Linguistica, one of the most widely-used unsupervised morphological parsers, by about 23%. [17]

Islam et al. (2007) proposed a fancy lightweight stemmer for Bangla and its use in spelling checker with a similar approach as proposed in [9]. The authors developed a suffix list and their proposed algorithm stripped the suffixes using the predetermined suffix list. It stripped a morphological variant based on the 'longest match' basis. They had collected a total of 72 suffixes for verbs, 22 for nouns and 8 for adjectives for Bangla language. The proposed stemming algorithm was primarily used for operating on inflections reducing derivationally related terms to the same stem. [18]

Yet Another Suffix Stripper (YASS) was a statistical approach developed by Majumder et al. (2007) [19]. The proposed algorithm used clustering techniques based on string distance measure. The authors claimed that it required no prior linguistic knowledge to improve recall of Information Retrieval (IR) systems against Indian languages. YASS was based on string distance measure. It managed to cluster a lexicon from a text corpus into homogeneous groups. Each group was expected to contain similar morphological variants of a single root word. They used Graph-theoretic clustering algorithm in their experiments.

III. Proposed Method for Bengali Stemming

Bangla words may contain multiple suffixes in a single stem. For example, the word "nirapottahInotai" contains three different suffixes 'hIno', 'ta', 'i'. In these cases, a 'longest suffix based' single stripping may not be sufficient. Therefore, we propose a recursive suffix stripping. We have developed a novel stemming technique based on three different approaches: conservative, aggressive, and rule-based. In conservative approach, we always ensure the stripped word to be in the predefined root list. However, aggressive approach always tries to eliminate all possible suffixes available in the word. The rule-based approach considers those suffixes whose identified patterns are different than the stripping patterns. Each of them has been elaborated in the following subsections. Before going into details of our algorithmic approach, we concentrate on the study of Bangla suffixes.

A. Suffix Study

Bangla is a highly inflectional language with many inflected forms of verbs, noun, adjective, and adverbs [18]. However, we have observed the suffixes from three different aspects. Some suffixes identify a specific class of words, like verbs. Some words contain pseudo-suffixes, which are not actually suffixes, rather they are part of the words. We have identified this type of suffixes and their stems or relative roots. We take these suffixes into consideration while working with conservative technique. We have considered 105 suffixes for our conservative approaches.

Moreover, we have observed a good variety of words, coming from noun origin or different language family contains 218 different suffixes for their variations due to articles, numbers, parts-of-speech conversion, and so on. These suffixes are used in aggressive approach of stemming.

Apart from this, there are a few suffixes, which have different nature than the suffixes described above. In case of these suffixes, we use a pattern in suffix identification, however, we hardly strip the whole suffixes. We strip quite differently from the word. For example, 'borrShay'. Here we identify 'ay', but we strip only 'y'. We use these suffixes in the rule-based approach. For the time being, we have used three different suffixes in this rule-based approach.

B. Recursive Suffix Stripping Algorithm

As we already described a word 'nirapottahInotai', which contains three suffixes. Some other word may contain a single, or two suffixes. In order to remove the undefined number of suffixes, we have introduced a recursive **Algorithm 1** to strip the suffixes repeatedly from the morphological variances to extract stem.

Algorithm 1: rsStripping($word$, $suffixes[]$, $stems$): An algorithm to strip all possible suffixes from a word

Input: $word$:to be stripped; $suffixes[]$: suffixes to be stripped; $stems$: vector of all possible stems
Output: Candidate stems list

```
1  if (word.length()<3) then
2  |  return;
3  for each suffix ∈ suffixes do
4  |  if (word.endsWith(suffix)) then
5  |  |  stem=word.strip(suffix)
6  |  |  if (stem.length()<2) then
7  |  |  |  continue;
8  |  |  if (lastChar(stem)==0x09CD) then
9  |  |  |  continue;
10 |  |  stems.add(stem)
11 |  |  rsStripping(stem, suffixes[], stems)
12 end
```

According to Recursive Suffix Stripping **Algorithm 1**, we take two data as input, ($word$ to be stripped and list of suffixes) and one reference address (call-by-reference) that contains a possible list of $stems$. The base criteria of the recursive function is stated at line 1. Then a simple iteration starts for all suffixes (refer to line 3). If the $word$ contains a suffix at the end (see line 4), algorithm strip it from the word to make a stem (line no. 5). However, if the length of the stem becomes less than 2 (as line 6), or the last character is 'hosonto' (refer to line 8) then the stem is not acceptable and starts processing for the next suffix. Otherwise, the stem is added to the stems vector (line no. 10) as a candidate stems and the function starts recursively for the stem word (see the line no 11).

C. Conservative Stemming

In conservative stemming, we use the Recursive Suffix Stripping **Algorithm 1** with the predefined list of suffixes. Then the **Algorithm 1** produces a vector or a list of candidate stems. Here, we only consider the stems, which are available in the root repository. This approach is called conservative as it only accepts the predefined set of forms (or roots). Among the short-listed stems, the stem with the smallest length is returned. Currently, we have 459 verb stems and 28 other stems in the root repository.

D. Aggressive Stemming

Aggressive stemming approach uses the same Recursive Suffix Stripping **Algorithm 1** with a different set of predefined suffixes. Similar to the conservative approach, the **Algorithm 1** produces a list of candidate stems. Unlike conservative approach, we consider all the stems and find the smallest length stems. In fact, the largest suffix has been removed aggressively. That is why, this approach is called as an aggressive approach.

E. Rule-based Stemming

Unlike conservative or aggressive approach, rule-based stemming is primarily applied for a separate list of suffixes. In reality, suffixes used in this list usually found at the end of a word. Therefore, we apply the first iteration of the algorithm as the rule-based approach, however, it becomes nothing but an aggressive approach after the first iteration. Evidently, rule-based stemming approach takes the essence of the aggressive approach.

Algorithm 2: ruleBasedStripping($word$, $rsuffixes[][2]$, $asuffixes[]$, $stems$): An algorithm to strip all possible suffixes from a word

Input: $word$:to be stripped; $asuffixes[]$: aggressive suffixes to be stripped; $rsuffixes[][2]$: suffixes for rule-based approach; $stems$: vector of all possible stems

Output: Candidate stems list

```
1  if (word.length()<3) then
2  |   return;
3  for each suffix-pair ∈ suffixes do
4  |   if (word.endsWith(suffix[i][0])) then
5  |   |   stem=word.strip(suffix[i][1])
6  |   |   if (stem.length()<2) then
7  |   |   |   continue;
8  |   |   if (lastChar(stem)==0x09CD) then
9  |   |   |   continue;
10 |   |   stems.add(stem)
11 |   |   rsStripping(stem, asuffixes[], stems)
12 end
```

According to Rule-based **Algorithm 2**, we take three data as input, ($word$ to be stripped, $rsuffixes[][2]$ be the suffixes of $n \times 2$ array for rule-based approach, and $asuffixes$ be a list of suffixes for aggressive approach) and one reference address (call-by-reference) that contains possible list of $stems$. The base criteria of the recursive function is stated at line 1. Then a simple iteration starts for each suffix-pair (refer to line 3). If the $word$ contains a suffix $suffix[i][0]$ at the end (see line 4), algorithm strip $suffix[i][1]$ from the word to make a stem (line no. 5). However, if the length of the stem becomes less than 2 (as line no. 6), or the last character is 'hosonto' (refer to line 8) then the stem is not acceptable, and starts processing for the next suffix-pair. Otherwise, the stem is added to the stems vector (line no. 10) as a candidate stems and the function calls Recursive Suffix Stripping **Algorithm 1** for aggressive suffixes that start recursion to find stem (see the line no. 11).

F. Proposed Stemming Algorithm

We have already discussed all of the components of our stemmer: the conservative approach, the aggressive approach, and the rule-based approach. Now, we put all the components together to extract a stem from the word. As an aggregation, we maintain three separate suffix-sets, $conservative\_suffixes[]$, $aggressive\_suffixes[]$ and $rule\_suffixes[][2]$ and $roots$ as a vector repository of some selected roots.

Algorithm 3: getStem ($word$): An algorithm to find stem of a given word

Input: $word$:to be stripped; $conservative\_suffixes[]$: suffixes usually associated with verb and dictionary words; $aggressive\_suffixes[]$: suffixes that can be associated with any words; $rule\_suffixes[][2]$: suffixes that have distinguished identification part and strip part; $stems$: vector of a stem(s); $stem$: a relative root of a word; $roots$: vector of selected roots

Output: Stem of a given word

```
1  rsStripping (word, conservative_suffixes[], stems);
2  stem=conservativeStemming(stems,roots);
3  if (stem!=null) then
4  |   return stem;
5  ruleBaedStripping(word, rule_suffixes[][2],
     aggressive_suffixes[], stems)
6  stem=aggressiveStemming(stems)
7  if (stem!=null) then
8  |   return stem;
9  rsStripping (word, aggressive_suffixes[], stems);
10 stem=aggressiveStemming(stem);
11 if (stem!=null) then
12 |   return stem;
13 else
14 |   return word;
```

According to the Recursive Suffix Stripping **Algorithm 1**, line no. 1 of the **Algorithm 3** produces a vector of stems. The function $conservativeStemming()$ at line 2 checks availability of every $stem \in stems$ in the $roots$. If it finds any match, it returns the smallest stem, otherwise returns null. Line 3 states the fact straight forward. It is like greedy method, out of the three approaches, if any earlier approach finds some solution, the rest is discarded (as line no. 3, 7, and 11). $ruleBasedStripping()$ function also generates a vector of stems (line no. 5). The role of the $aggressiveStemming()$ is nothing but finding the smallest stem in $stems$ blindly. Therefore, the smallest stem is extracted in line 6. However, if the $stems$ is empty, then stem must be empty or null, which means that the $ruleBasedStripping()$ is unsuccessful in finding any stem. According to line 9, the $rsStripping()$ function is applied to generate a list of candidate $stems$, then the $aggressiveStemming()$ produces the smallest stem unless vector $stems$ is not empty. If all the steps above fail, the algorithm returns the original word as $stem$ as stated in line 14.

IV. Experiments and Evaluation

To evaluate the effectiveness of our proposed Bengali stemmer, we create a larger dataset by crawling news article from a popular Bengali daily news paper, "The Daily Protho Alo" [1]. The dataset contains almost 0.78 million words with 62 thousand distinct words. Our stemmer reduced this 62 thousand words into 32 thousand stems.

TABLE I. Basic Statistics about Dataset

Topic	Quantity
Number of documents	3772
Number of total sentences	64972
Number of distinct sentences	58058
Number of total words	779064
Number of distinct words	62367
Number of extracted stems	32074

In Information Access (IA) research, a document is considered as a vector of distinct words. Therefore, our proposed stemmer certainly reduces the vector dimensions by an almost half, which may increase the performance twice faster. Moreover, among the 31074 extracted stems 29479 stems are closely accurate, which means that almost 92% extracted stems are accurate with a minimal root repository. This experimental result shows its strength and efficiency against other available systems.

V. Conclusion and Future Direction

In this paper, we proposed an efficient and effective method for Bengali stemming based on three different approaches (conservative, aggressive, and rule-based) during stemming. Firstly, conservative stemming addresses mainly verbs and exceptional words heavily affected by stemmer. In this technique, 459 verb roots and 28 exceptional word roots are collected to form a small repository. The conservative stemming approach only considers the stems available in the repository. As the name suggest, the fundamental criteria of the conservative approach are that it does not allow stems beyond the predefined root-set. In the aggressive stemming approach, suffixes are eliminated repeatedly until the last suffix being removed. As the name suggests it finds the smallest length stems through eliminating the largest suffix from the word aggressively. Other than these two techniques, rule-based approach focuses on some right most suffixes, where stripping suffixes are different than the identified patterns. Rule-based approaches is a bit modified aggressive approach. With the essence of these three techniques, our novel stemming algorithm becomes unique and efficient for Bangla. We have experimented against a relatively larger dataset containing 0.78 million words in a 3772 real news documents dataset. The dataset contains heterogeneous words, of which many have been imported from different languages. In spite of the heterogeneous nature of our dataset, we achieved 92% of accuracy. However, because of the aggressive approach, we have observed several over stemming, which can be our future direction. Moreover, we will also focus on more purified suffixes and enlarged root repository containing most of the exceptional words.

[1] http://www.prothom-alo.com

Appendix

Our stemmer can be downloaded from our laboratory page http://skeim.org/projects-download/. It is a runnable *.jar. Anyone can utilize this software not only his Information Access (IA) research project, but also make use of it as an application. The short description on operation has been depicted in Table II. The basic command reflects:

$$java\ -jar\ jSkeimStemmer.jar\ <parameter>\ \ <data>$$

The basic operations has been articulated in Table II.

References

[1] J. Lovins, "Development of a stemming algorithm," *Translation and Computational Linguistcs*, vol. 11, no. 1, pp. 22–31, 1968.

[2] M. F. Porter, "An algorithm for suffix stripping," *Program*, vol. 14, no. 3, pp. 130–137, 1980.

[3] D. Hull *et al.*, "Stemming algorithms: A case study for detailed evaluation," *JASIS*, vol. 47, no. 1, pp. 70–84, 1996.

[4] W. Kraaij and R. Pohlmann, "Viewing stemming as recall enhancement," in *Proceedings of the 19th annual international ACM SIGIR conference on Research and development in information retrieval*. ACM, 1996, pp. 40–48.

[5] M. Adriani, J. Asian, B. Nazief, S. Tahaghoghi, and H. Williams, "Stemming indonesian: A confix-stripping approach," *ACM Transactions on Asian Language Information Processing (TALIP)*, vol. 6, no. 4, pp. 1–33, 2007.

[6] V. Gupta and G. Lehal, "Punjabi language stemmer for nouns and proper names," in *proceedings of the 2nd Workshop on South and Southeast Asian Natural Language Processing (WSSANLP) IJCNLP*, 2011, pp. 35–39.

[7] S. Tai, C. Ong, and N. Abullah, "On designing an automated malaysian stemmer for the malay language (poster session)," in *Proceedings of the fifth international workshop on on Information retrieval with Asian languages*. ACM, 2000, pp. 207–208.

[8] V. Gupta and G. Lehal, "A survey of common stemming techniques and existing stemmers for indian languages," *Journal of Emerging Technologies in Web Intelligence*, vol. 5, no. 2, pp. 157–161, 2013.

[9] A. Ramanathan and D. Rao, "A lightweight stemmer for hindi," in *Proceedings of Workshop on Computational Linguistics for South-Asian Languages, EACL*, 2003.

[10] M. Greengrass, A. Robertson, S. Robyn, and P. Willett, "Processing morphological variants in searches of latin text," *Information research news*, vol. 6, no. 4, pp. 2–5, 1996.

[11] S. Bressan, "Indexing the indonesian web: Language identification and miscellaneous issues," *Recall*, vol. 80, p. 100, 2000.

[12] J. Carlberger, H. Dalianis, M. Hassel, and O. Knutsson, "Improving precision in information retrieval for swedish using stemming," in *Proceedings of NODALIDA 13th Nordic Conference on Computational Linguistics*, 2001, pp. 21–22.

TABLE II. THE BASIC OPERATION MANUAL

Parameter	Data	Output	Comments
-w	nirapottahInotai	nirapotta	Only the final stem has been returned as output.
-wa	nirapottahInotai	The first [ ] contains candidates of conservative stemming, if appropriate stem is not found. Second [ ] shows the list of candidate stems from aggressive approach.	This can be used for analysis purpose.
-f	name of the file which contains Bangla text	The first line contains the original text, whereas the second line represent the one by one corresponding stemmed text.	The final stem of each word has been returned sequentially as output.
-fa	name of the file which contains Bangla text	Full analysis of each word available in the file as described in row 2.	This can be used for analysis purpose.

[13] C. Monz and M. De Rijke, "Shallow morphological analysis in monolingual information retrieval for dutch, german, and italian," in *Evaluation of Cross-Language Information Retrieval Systems*. Springer, 2001, pp. 262–277.

[14] I. Moulinier, J. A. McCulloh, and E. Lund, "West group at clef 2000: Non-english monolingual retrieval," in *Cross-Language Information Retrieval and Evaluation*. Springer, 2000, pp. 253–260.

[15] M. Popovic and P. Willett, "The effectiveness of stemming for natural-language access to slovene textual data," *Journal of the American Society for Information Science*, vol. 43, no. 5, p. 384, 1992.

[16] F. Ekmekcioglu, M. Lynch, and P. Willett, "Stemming and n-gram matching for term conflation in turkish texts," *Information Research News*, vol. 7, no. 1, pp. 2–6, 1996.

[17] S. Dasgupta and V. Ng, "Unsupervised morphological parsing of bengali," *Language Resources and Evaluation*, vol. 40, no. 3-4, pp. 311–330, 2006.

[18] M. Islam, M. Uddin, and M. Khan, "A light weight stemmer for bengali and its use in spelling checker," in *Proceedings of the 1st International Conference on Digital Communication and Computer Applications (DCCA07), Irbid, Jordan*. BRAC University, 2007.

[19] P. Majumder, M. Mitra, S. Parui, G. Kole, P. Mitra, and K. Datta, "Yass: Yet another suffix stripper," *ACM transactions on information systems (TOIS)*, vol. 25, no. 4, p. 18, 2007.

Towards Reducing BoP Penalty through Rural E-Commerce: Optimization of Product Delivery Mechanism

Kazi Mozaher Hossein
Dept. of Advanced Information Technology
Kyushu University, Fukuoka, Japan
kmhjewel@f.ait.kyushu-u.ac.jp

Md. Nazmul Hossain
Dept. of Advanced Information Technology
Kyushu University, Fukuoka, Japan
nazmul@f.ait.kyushu-u.ac.jp

Rajib Chakrabarty
Global Communication Center
Grameen Communications, Dhaka, Bangladesh
rajib@mail.grameen.com

Fumihiko Yokota
Institute of Decision Science for a Sustainable Society
Kyushu University, Fukuoka, Japan
yokota.fumihiko.785@m.kyushu-u.ac.jp

Hironobu Kitaoka
Toyota Motor Corporation, Aichi, Japan
hironobu_kitaoka_aa@mail.toyota.co.jp

Hiroshi Okajima
Toyota Motor Corporation, Aichi, Japan
hiroshi_okajima_aa@mail.toyota.co.jp

Ashir Ahmed
Dept. of Advanced Information Technology
Kyushu University, Fukuoka, Japan
ashir@ait.kyushu-u.ac.jp

***Abstract*—** E-Commerce became popular among the affluent people in the world. However, a big portion of the population of the world cannot enjoy the advantages of e-commerce service because they do not have (1) access to online catalog (2) payment system to pay for online purchase (3) home delivery infrastructure in their community. This scenario ultimately increases the BoP penalty (people at Base of the Economic Pyramid pays more than the urban affluent people for the same product). This article introduces the existing e-commerce operation model and outline the barriers that limit its expansion to cover the underserved community. The recent penetration of 3G technologies and smart phones allow the villagers to access to internet based e-commerce services. Availability of mobile money transfer service made it easier for the villagers to make remote payments. A significant number of e-commerce start-ups opened new windows for people in the urban areas to enjoy e-commerce services. The remaining biggest hurdle for the villagers is now the delivery mechanism in the rural areas in a cost effective way. We take this challenge to reduce BoP penalty by delivering goods in a cost-effective way. In this article, we propose a model to deliver the ordered product to a nearby service point. The model studies the required products in the village and their demand frequencies. Our approaches are 1) Collecting group order and 2) Efficient delivery mechanism. We have two experimental sites (population more than 0.3 million people in each site) in Bangladesh. We have simulated our model with a social good and found that 112.5% BoP penalty can be reduced in an ideal situation.

Keywords—BoP Penalty; ICT; E-commerce; Last mile delivery; GramWeb

I. INTRODUCTION

Currently, there are 4 billion people at the BoP (Base of the Pyramid), comprising 69% [1][2] of the world population. Despite their low income and limited purchase capacity, they make frequent purchases within their limited spending power. The people at the BoP often pay higher prices for basic goods and services than do wealthier consumers [3]. Considering the definition of BoP penalty we can draw following equations:

$$\text{BoP penalty} = \text{Premium price} - \text{Original price} \ldots\ldots..(1)$$

$$\text{Poverty premium} = \frac{\textit{Premium price}}{\textit{Original price}} \text{ times} \ldots\ldots\ldots\ldots(2)$$

Original price of a product or service is the price determined by the seller of that product or service in a competitive marketplace. On the other hand Premium price is the increased price of the same product or service in a remote area due to extra delivery cost added with the original price of the product or service. From equation 1 we find BoP penalty is the difference between premium price and original price which is measured in monetary unit. In equation 2 poverty premium is the degree of price difference. For example, if a product's price in a competitive market place is 10 dollars and the price of that product in a remote area is 15 dollars then poverty premium would be 1.5 times and BoP penalty would be 5 dollars. Table 1 shows some practical examples of BoP penalty and poverty premium in Bangladesh perspectives.

Table 1: Examples of BoP Penalty

Item	Bheramara (Rural)	Dhaka (Urban)	Poverty Premium (times)
Mineral water (500 ml)	20 taka	15 taka	1.33
Blood glucose test	300 taka	150 taka	2.00
Rice (per kg)	60 taka	50 taka	1.20

Cost disparities between BoP consumers and the rich in the same economy can be explained only by the fact that the poverty penalty at the BoP is a result of inefficiencies in access to distribution and the role of the local intermediaries [4]. Table 1 shows three examples that describe how people in a rural area buy same products or services in higher prices than the people live in an urban area.

Rural E-commerce can be used as a tool to reduce BoP penalty. It will enable villagers to purchase products from a village at any time of the day and get the desired products delivered to their doors. It saves time, money and labor [5]. A product seller can upload the product information on the web and can breach the boundaries of the local market to reach the customers on a global scale. A customer, on the other hand, can search for the desired product in a much more extensive selection space, and find a suitable product. In this way, e-commerce brings benefits for both the buyers and sellers as indicated by the trend in e-sales.

In order to purchase a product through a web-based e-commerce service, a customer needs access to the internet and an online payment mechanism, typically a credit card. Presently 40% of world population has access to the internet [6]. Fig. 1 shows a big gap between the users of internet in developed world and developing world. Four billion people from developing countries remain offline, representing 2/3rd of the population residing in developing countries [7]. Out of 940 million people living in the least developed countries (LDCs), only 89 million use the internet, corresponding to a 9.5% penetration rate [7]. Similarly, most of the people of developing and least developing countries are unable to use credit cards due to the rules and regulation which are also designed targeting the rich people of our society. However, mobile money transfer is becoming popular to the local e-commerce companies which is replacing the need of credit card in the developing and least developing countries. Still, product delivery in a cost-effective way is the biggest challenge at the BoP community.

Toyota Motor Corporation and Kyushu University in Japan are jointly carrying out a research to create values by

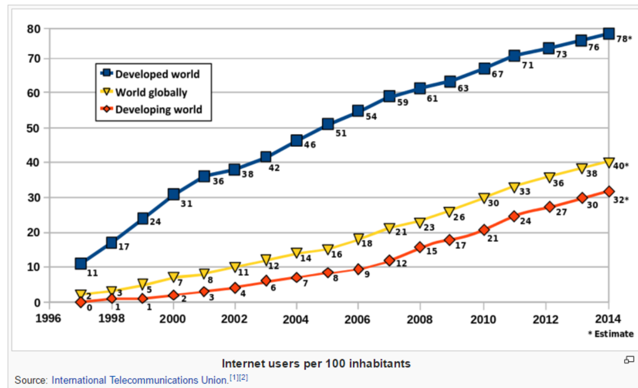


Figure 1: Internet Users Worldwide

introducing two community cars in two experimental sites. Grameen Communications in Bangladesh is supporting the experiment. A 10 seated Toyota Hiace vehicle carries 4 major services (healthcare, education, learning and purchase), we call it SSW (Social Services on Wheels) project [8]. In this research, we use this experimental platform to verify our rural e-commerce model. Our aim is to reduce BoP penalty. We design an ICT based product delivery mechanism and deliver social goods. By social goods, we mean the products that are required in the community, can solve a social problem and can create a big social impact. People may be unaware of the product or the purchase volume is not big enough to justify financial benefits to attract a local seller. Examples of these social products are: malaria preventive mosquito nets, energy efficient bulbs, sanitary napkins for females, medicines etc. These products are available in Dhaka city but not widely introduced or sold in local markets. In our proposed e-commerce system, we develop a villager-friendly online catalog to list the basic products and update the catalog based on villagers' needs.

Two experimental sites are in two districts in Bangladesh. One is in Kalhitari Upazila under Tangail district (105 km away from the capital Dhaka city) and the other one is in Bheramara Upazila under Kushtia district (235 km away from the capital Dhaka city).

Table 2: Demographic information of experimental sites

	Site-1 (Kalihati)	Site-2 (Bheramara)
Population (million)	0.41	0.2
Area (sq. km.)	295.6	153.7
Population Density (sq. km.)	1388	1302
Distance from Dhaka City (km.)	105	235
Courier delivery service point (km.)	8 to 18	2 to 14
Number of mobile phones (million)*	0.18	0.09
Percentage of people know about e-commerce	Under survey	Under survey
Availability of Smart Phones (million)*	0.0144	0.0072
Percentage of young (10-24 years) population (million)*	0.13	0.062

*Estimated

We have identified the following reasons for the increased BoP penalty- (1) poor distribution network (2) low frequency of demand (3) unnecessary intermediaries in the supply channel. We argue that the penalty can be reduced completely or partially if the following can be implemented- (1) group purchase (2) group delivery. Group purchase and group delivery will increase the delay but will reduce the total cost of the product.

Section II introduces the traditional e-commerce system wherein the participation of the local villagers are missing. Section III describes the obstacle of delivery model where the BoP penalty affects the most. Section IV explains our proposed model. We came up with a scheduling mechanism for group purchase, group distribution model in order to reduce the product cost i.e. BoP penalty. We summarize our findings in section V.

II. TRADITIONAL E-COMMERCE SYSTEM

We study traditional e-commerce system to understand most

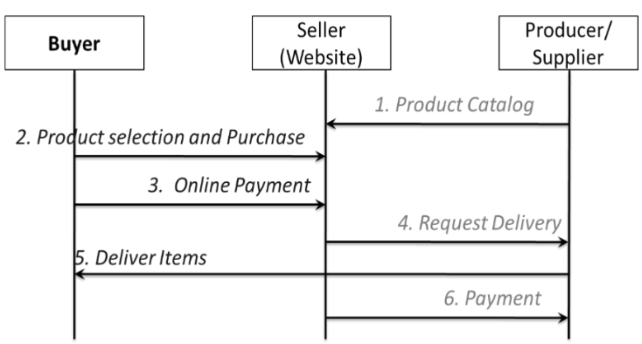


Figure 2: Traditional E-commerce System

important e-commerce components and major stakeholders of e-commerce. Fig. 2 shows a typical e-commerce model. There are three major stakeholders in the present e-commerce system:

Buyer: In a web-based e-commerce system, a buyer visits the product website, selects the product, and makes the payment by credit card, PayPal or other online transaction system.

Seller: A seller is a website that interfaces with buyers. Such a website offers a product catalog, an interactive interface to receive customers' preferences, a shopping cart system for the purchasing process and an online payment system. The website needs to ensure a secure system to handle customer information.

Supplier: A supplier receives purchase orders through the web, checks the payment process, and ensures the delivery of the product and post-sales services. A supplier can be a producer of the product, its representative, an agent or a distributor.

There are three components of e-commerce essential to operate its function by the stakeholders. They are:

a) Internet accessibility to explore online product catalog

b) A payment mechanism such as credit card, mobile money transfer etc.

c) Home delivery mechanism to transport the purchased product from supplier's point to buyer's point.

Therefore, the model can work only in places where the basic infrastructure of e-commerce is ready. Unfortunately, the remote locations and low income people are not included in the customer list of the current e-commerce system. However, recent internet penetration and popularity of mobile money transfer in developing countries are triggering its popularity in the rural areas. E-commerce service lies incomplete for the rural people of developing countries due to lack of proper home delivery mechanism. In the next section, we explain current product delivery scenario for rural e-commerce in Bangladesh.

III. THE BIG OBSTACLE

In Bangladesh, there is an established product distribution channel among city areas of different districts through

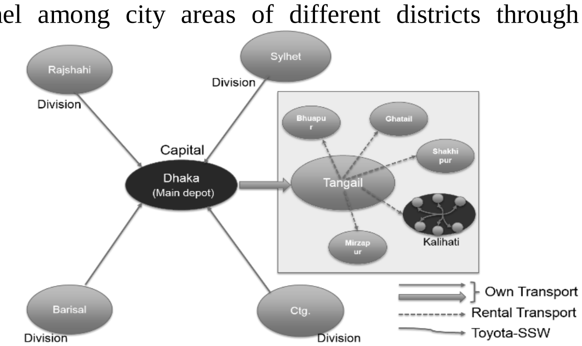


Figure 3: Existing Product Delivery Mechanism

different courier service companies. E-commerce service is also getting popular in those areas. From the city areas products are delivered to the sub-district branches using rental transport facility but still with the responsibility of the courier service companies. However, there is a big gap of delivering products from the sub-district office to the rural households. From Fig. 3 we find using the courier service channel, products are transporting to Dhaka (central depot.) from divisional level. From Dhaka, the products for Tangail district are sorted and transported. From Tangail, products are again sorted and transported to Kalihati, a subdistrict of Tangail. From Kalihati, social goods are delivered to the rural service points using the GramCar of Toyota-SSW project. In next chapter, we discuss in-detail about social goods, GramCar and Toyota-SSW project.

IV. Our Approach: Scheduling Operation Towards Maximum Utilization of Resources

In order to reduce the BoP penalty, we have identified the following items – (1) number of products to be ordered from the same area (2) distance from the existing courier service point.

For our pilot study, we used the research platform of GramCar, a joint research project of Kyushu University and Toyota with the help of Grameen Communications in Bangladesh. They are carrying out an experiment in Bangladesh to evaluate whether a vehicle can carry multiple services to the villagers' doorstep. The social services are healthcare service (e-health, virtual blood bank), mobility service (college bus, emergency transport, and demand responsive transport), education service (e-learning) and social goods delivery. At the rural sites the village car entrepreneur (VCE) operates the services with direct support of urban head office including doctors call center, database management, software development and maintenance, social goods collection and supply, training of rural staffs, promotional design and all other logistic support.

Under social goods delivery service a community car carries social good samples, Wi-Fi tablets and SSW ICT trainer to the service point. Social goods include, sanitary napkins and underwear garments, energy efficient bulbs, solar torches and charges, and mosquito nets and repellents. The social good samples allows the villager's to see the purchasable items. The villager and SSW ICT trainer use a catalog to access online purchasing opportunities. To overcome any illiteracy issue, a catalogue with pictures and prices are shown to the villagers. The customer selects which items he or she wishes to purchase and the ICT trainer makes the order online on the customer's behalf. The ordered products are delivered to the service point one week later by the community car.

A. Shared Resource, Shared Cost

In our model sites, the GramCar visits different service points in the community in a scheduled basis. The service points are scheduled in such a way as one service point serves the rural people once a week. Utilizing the empty space of the GramCar the social goods are delivered once a week in a specific service point without spending any extra money on last mile product delivery.

B. Product Selection

We select products based on the demand of rural community that adjusts with our delivery mechanism. Perishable products are excluded from the product list as the products are delivered one week after order placement. Products are also selected which have the attribute of solving rural

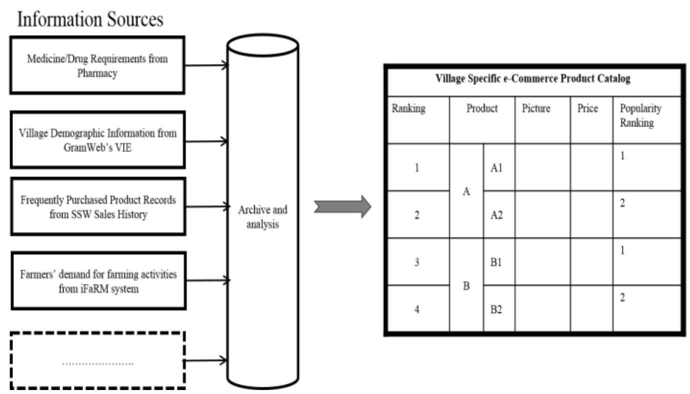


Figure 4: Model of a Community Specific Online Catalog

problems. Locally unavailable products are included in the product list to save money, time, and effort of the rural people. Identifying community demand is an important task. We are analyzing different online/offline activities of rural people such as healthcare receiving, movement for family shopping, farming activities and so on. We are analyzing the data that we receive from ICT based social services as well as the data from different offline sources such as questionnaire survey and observation. Based on the analysis we carefully include new products in the product catalog. Fig. 4 shows a model of community specific online catalog. It will be prepared after the completion of community data analysis. The catalog will be displayed in GramWeb platform. GramWeb is an information platform for low-literate people that connects each and every villages of Bangladesh.

C. Service Point Optimization

We plan and implement different service point models to design an appropriate product delivery mechanism for the rural community. Fig. 5 shows the model of bi-weekly and weekly service points which are already applied in the experimental sites. In the bi-weekly model the vehicle visits 10 service points in the community per month. One service point deliver social services once in every two weeks.

On the other hand, under the weekly service point model the vehicle visits and deliver social services to 5 service points per month. One service point is repeated once a week. Findings from our experiment on the above mentioned weekly and bi-weekly service point models are the bi-weekly service point model serves more new people than weekly service point model while weekly service point model serves more repetitive customers.

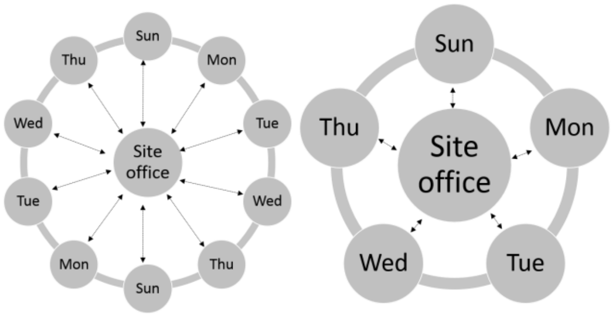


Figure 5: Service Point Design (Bi-Weekly and Weekly)

In order to increase service frequency in the rural community we are working on a new service point model (Model-3). Under this model the vehicle will touch 6 service points in one day. There will be total 30 service points and every service point will be repeated after one week.

Table-3 shows a comparison chart among different service point models.

Table 3: Different Social Goods Delivery Service Point Models

Measuring items	Model-1 (weekly)	Model-2 (bi-weekly)	Model-3 (proposed)
Visiting frequency (per month per service point)	4 times	2 times	4 times
Total service points in one site	5 service points	10 service points	30 service points
Total customer reach (per Month)	1000 households	2000 households	6000 households
Maximum product delivery time required	7 days	14 days	7 days

D. Social Goods Delivery Operation

Current social good delivery operation is shown in fig. 6. The process starts from customer order placement. The village car entrepreneur (VCE) transfers the order information to the head office. The supplier receives the product request from the head office. Product delivery is taken place in an opposite direction. The payment flow is similar to the order flow which means it starts from the customer and ends at the supplier. Mobile money transfer technology is currently using for the payment mechanism.

E. Group Delivery to Reduce BoP Penalty

Aschorjo moshari, a long lasting insecticide-treated mosquito net, is now available in Bangladesh to protect from a mosquito borne disease malaria. The customers can buy one mosquito net from the urban market at 850 Bangladeshi

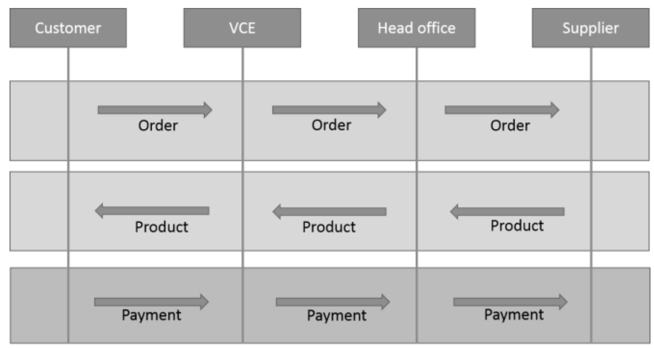


Figure 6: Social Goods Delivery Operation

Taka. If the product is ordered from the remote area of rural Bangladesh through e-commerce site it will add additional 100 Taka and will be delivered until the sub-district level. They have to travel all the way through uncomfortable transport which will add additional money, time and labor. We have selected this product to deliver at the remote area of rural Bangladesh. We receive a group order from the rural entrepreneur. We collect the product from the supplier at Dhaka with a wholesale price 625 Taka. We send through existing courier service. The courier company delivers the package up to the sub-district point and charge 200 Taka for 10 mosquito nets. SSW-GramCar collects the mosquito nets and distribute it at the rural service points. The rural consumers who ordered the product previously collect it from their nearest service point.

By analyzing customer demand and collecting group order we can buy the mosquito net from the supplier at wholesale price and can send them in reduced delivery cost by distributing the cost among multiple products. Fig. 7 shows total delivery cost for 1 mosquito net from supplier to the service point is 38 Taka only. We sell one mosquito net at 750 Taka. From this experiment we find the rural consumers are able to buy the mosquito net by reducing the BoP penalty more than 200 Taka with reduced time and effort as well.

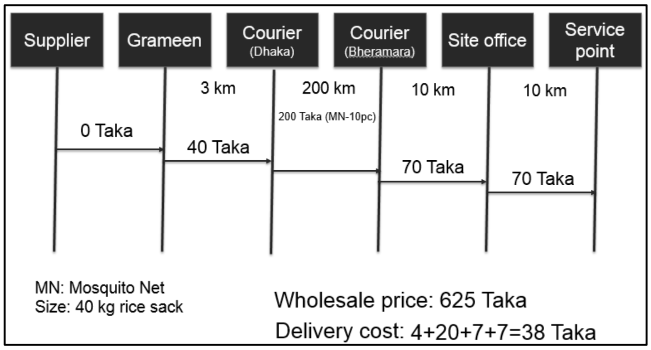


Figure 7: Financial Projection of Product Delivery

V. SUMMARY AND CHALLENGES

The article focused on the issue of BoP penalty. This article showed a way to reduce the BoP penalty by improving the prodoct delivery mechanism. We focused on the efficient product delivery mechanism and last mile delivery problem and came up with group delivery methotd after analyzing community demand. We simulated the model in two rural areas in Bangladesh to verify our model. Considering equation 1 of measuring BoP penalty and we found that BoP penalty for mosquito net is -100 taka which means neither the BoP penalty is reduced partially nor fully, rather we could deliver the mosquito net at a lower price than the urban marketplace price. The BoP penalty was reduced by 112.5%.

The challenges we faced in delivering social goods outlined below:

- Individual purchase is more popular than group purchase. Due to lack of promotion and/or source of information people are less aware about the benefit of group purchase.

- Many people do not know about the advantages and the risks of e-commerce system. In the country, there are lot of fake items on e-commerce sites and there is no good platform to take care of post sales claims.

- Not all the products can be sold through e-commerce systems. Perishable products will not be popular until we have a safe and secure delivery and storage mechanism.

- Mobile money transfer is not secure yet. Technology needs to be improved to gain the trust of the people for mobile money transfer.

- Traditionally people are not accustomed with the culture of pre-paid system for unseen products.

- The online catalog needs to be villager friendly. Consumers trust on buying certain goods that they can't touch previously. Rural people checks the quality of the product by touching, tasting or taking closer look. The online catalog does not have a product before making a purchase decision. But in e-commerce shopping such kind of option is absent.

  Our next step will concentrate on the following items.

- Prepare a community specific online catalog based on community demand.

- Evaluate the new model with increased service points.

- Outline the requirements to establish strategic alliance with new suppliers based on the demand of rural people.

- Estimate the reduced BoP penalty for newly added products.

REFERENCES

[1] T. London, R. Anupindi, and S. Sheth, "Lessons learned from ventures serving base of the pyramid producers," Journal of Business Research 63, pp. 582–594, 2010.

[2] A. Hussain, "E-commerce and beyond: opportunities for developing country SMEs," http://tradeforum.org/article, December 2013.

[3] A. Hammond, W. J. Kramer, J. Tran, R. Katz, and C. Walker, "The Next 4 Billion: Market Size and Business Strategy at the Base of the Pyramid", World Resources Institute Publications, March 2007.

[4] C. K. Prahalad, "Fortune at the Bottom of the Pyramid", New Jersey: Pearson Publishing, 2005.

[5] A. Ahmed, M. A. Rahman, and T. Ohsugi, "E-commerce for the unreached community," Proc. IADIS International Conference ICT, Society and Human Beings (ICT 2011), July 24-26, 2011, Rome, Italy.

[6] International Telecommunication Union (ITU), "Individuals using the Internet 2005 to 2014", Key ICT indicators for developed and developing countries and the world (totals and penetration rates)", May, 2015.

[7] International Telecommunication Union (ITU), "ICT Facts and Figures: The World in 2015", May 2015, Geneva, Switzerland.

[8] J. Kamau, A. Reberio-Hargrave, H. Saito, E. Abdullah, H. Okajima, and A. Ahmed, "Social Services on Wheels: A sustainable model to improve access in unreached communitites", IST-Africa roceedings, International Information Management Corporation IIMC, 2014

[9] R. Varadarajan, "Fortune at the bottom of the innovation pyramid: The strategic logic of incremental innovations," Business Horizons 52, pp. 21-29, 2009.

[10] M. Rivera-Santos, and C. Rufin, "Global village vs. small town: understanding networks at the base of the pyramid," International Business Review 19, pp. 126-139, 2010.

[11] S. Gold, R. Hahn, and S. Seuring, "Sustainable supply chain management in 'Base of the Pyramid' food projects-A path to triple bottom line approaches for multinationals?", International Business Review 22, pp. 784-799, 2013.

[12] S. Rashid, "JITA Social Business Bangladesh Limited-Reaching the unreachable in Bangladesh", Field Actions Science Reports, Special Issue 12, October 2014.

[13] R. Rahayu, and J. Day, "Determinant Factors of E-commerce Adoption by SMEs in Developing Country: Evidence from

Indonesia," World Conference on Technology, Innovation and Entrepreneurship, Procedia - Social and Behavioral Sciences 195 (2015) 142 – 150.

[14] S. Kurnia, J. Choudrie, R. M. Mahbubur, and B. Alzougool, "E-commerce technology adoption: A Malaysian grocery SME retail sector study," Journal of Business Research, 68 (2015), 1906-1918.

[15] S. H., Liao, C. M., Chen, and C. H., Wu, "Mining customer knowledge for product line and brand extension in retailing," Expert Systems with Applications 34 (2008) 1763–1776.

[16] N. Kshetri, "Barriers to e-commerce and competitive business models in developing countries: A case study," Electronic Commerce Research and Applications 6 (2007) 443–452.

[17] Tsuboi, H., Mosquito Nets as Social Business in Bangladesh — A Case Study of BASF Grameen Ltd., Akita University Graduate School of Engineering and Resource Studies Research Report, Japan, vol. 34, October, 2013.

[18] A. Ahmed, L. Kabir, and H. Yasuura, "An information platform for low –literate villagers," Proc. IEEE 24th International Conference for Advanced Information Networking and Applications (AINA 2010), April 20-23, Perth, Australia.

ECG Signal Based Heart Disease Detection System for Telemedicine Application

[1]Emranul Haque and [2]Feroz Ahmed

Department of Electrical and Electronic Engineering
Independent University, Bangladesh

Dhaka, Bangladesh

[1]emran1612@gmail.com and [2]fahmed@iub.edu.bd

Abstract— **Most of the cardiac disorders are arise due to irregular rhythm of the heart. Irregular heartbeats leads to abnormal PQRST values which can be traced from patient's ECG. This paper presents an automatic ECG signal processing system for detecting cardiac disorders as part of telemedicine application. The proposed system acquires ECG signal, processes it and extracts important parameters like PQRST to detect heart abnormalities. This system also includes an integrated remote transmission module by which recorded data can be transmitted over the internet to the central healthcare database system. The proposed system has been designed using Labview application software, MATLAB and C programming language and tested on ECG database obtained from MIT-BIH and synthetic ECG generated by NI DAQ device.**

Index Terms*— *Heart disease; ECG; Virtual instrument; G-server; QRS complex; Telemedicine; DAQ.

I. INTRODUCTION

Predicting and diagnosing heart diseases such as sudden cardiac death (SAD), pulmonary diseases, AV block etc through ECG has become one of the major topics in the field of medical science and biomedical engineering. Electrical activity of the heart can be traced using ECG which appears as periodic signal. A complete ECG cycle is marked as P wave, QRS complex, T wave and sometimes U wave. Diagnosis is performed based on these extracted features. So, precise and accurate extraction of ECG signal is always essential as it provides invaluable information to the physician that helps them reach at better diagnosis decision and treatment for the patients.

According to medical terms, most of the important information can be derived from the P wave, QRS complex and T wave. These parameters are divided into PR interval, PR segment, QRS interval and ST segment as shown in figure 1. However, due to baseline wandering, power line interference, noise and amplitude of T wave similar to QRS complex create problems to separate each of this parameter properly and accurately.

Several research activities have been dealing with the detection of the above mentioned ECG parameters. Various kinds of special digital filters [1-4] have been proposed to detect ECG features. Pan and Tomkin's proposed [5] algorithms analyze peak value to recognize QRS complex. Multiscale Wavelet Transforms [2-4] also proposed by many researchers. First derivative based Slope Vector Waveform (SVM) [7] was proposed to detect the complex points. Pattern or template based methods [6] are discussed for detection of R point. All of the proposed method [8-13,28] requires large mathematical calculation and thus computational overhead became an issue for these proposed methods.

The concept of remote healthcare system or Telemedicine has been introduced in the last few decades. Telemedicine system has gained a huge attention due to its huge potential and application [15]. As a result, the number of implemented telemedicine system has been increasing [15-17]. Generally, this system takes the patients data sent it to the central system. The central system processes the data to detect diseases (if any). A real time monitoring option is also included in some system [18-22, 25-27]. Central health system needs a good amount of time to process huge data of different patients and provide report to each patient. Real time monitoring is only available when the physician is present on the other side of the system. This scenario may increase the suffering of the patient especially with cardiac disorder if not detected at an initial stage. Detection of diseases within the system itself based on patients data can be a possible solution of the problem mentioned above.

In this paper, fast and simple techniques have been employed to extract ECG parameters and detect cardiac disorder. First simple digital filters are used to remove noise. Then Wavelet analysis has been applied for further de-noise the signal and remove baseline wandering. From the clean ECG, it is easy to extract important ECG parameters using threshold method. All of the signal processing has been done using Labview in conjunction with MATLAB [2] and C which makes the proposed system less computationally intensive compared to the others. After successfully extract the ECG parameters, a comparison is done with the normal ECG parameters. Any mismatch will indicate abnormality in the current ECG. In addition, the proposed system also includes a web based telemedicine module which can be used for real time doctor-patient interaction and also for sending the recorded data to the central remote station for further verification.

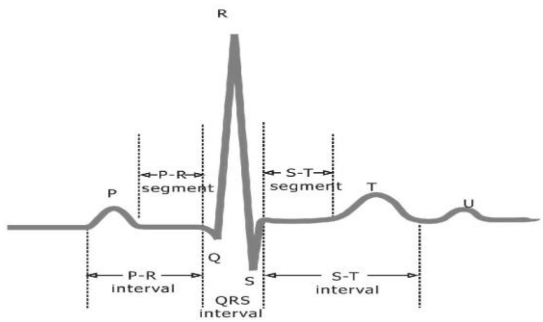


Fig 1. Different segments of ECG [24]

II. Structure of the Proposed System

The general structure of the proposed system is shown is figure 2. The system is divided into four major segments which are described here:

1) Data acquisition and generation module: This module takes ECG data from different sources. Sources include real time ECG data from the electrode through DAQ and from different ECG databases. In this work, ECG data are taken from the MIT-BIH[30] database and used this data to generate ECG signal through the analog output of the DAQ. The purpose of the generating ECG signal through DAQ is to include the effect of noise and interference of different sources in the signal.

2) Pre-Processing Unit: This module consists of lowpass and bandpass digital filters. These filters are used to remove the noise which is added to the signal. After that Wavelet analysis is applied to make the signal noise free which is shown in figure 3 [a,b].

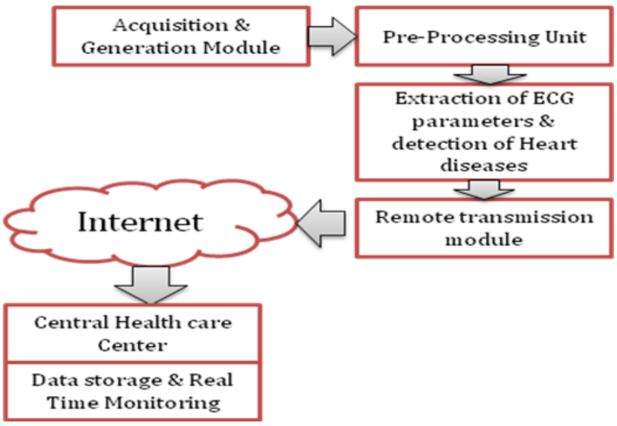
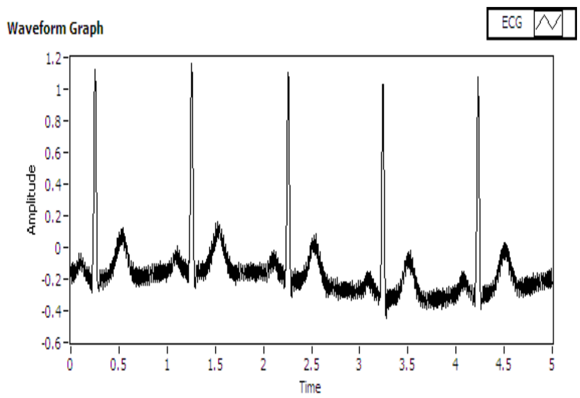
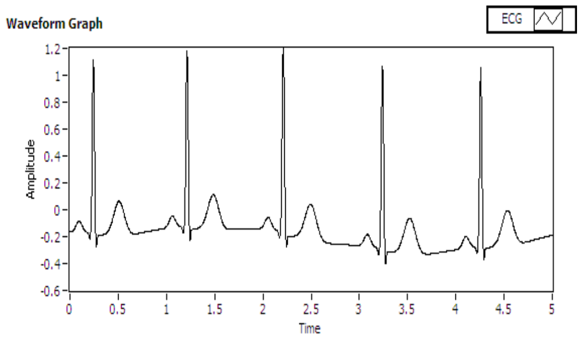


Fig 2. Structure of the proposed system

3) ECG features Extraction unit: When the noise free signal enters this module, it started detecting peaks/valleys [21-22] of the signal. Based on threshold method this module extract P,Q,R,S and T points. This unit also calculates R-R interval to calculate heart beat rate. For each ECG wave the threshold values are given in the table 1 and shown in figure 4.



(a)



(b)

Fig 3. a) Noisy ECG b) Filtered ECG

TABLE I

Threshold values

ECG parameter	Amplitude(mV)
P-wave	0.25
R-wave	0.8-0.12
Q-wave	25% of R wave
T-wave	0.1-0.5

4) Web based communication and transmission module: The unit utilizes the G-server which is included in the labview application software. With the G- server the recorded data can be sent to central system or database. In critical situation where emergency steps are required, this module can provide a real time doctor-patient communication interface.

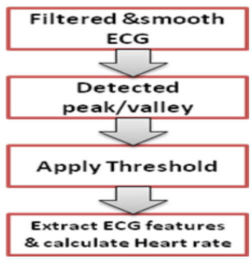


Fig 4. ECG feature extraction method

III. Detection of Heart Abnormalities

Any changes in ECG parameters from their normal values reflect cardiac disorders. For example, any elongation in PQ segment and QT interval indicate heart block and congenital disorders. After the extraction of ECG parameters, this system compares the values with the predefined normal values and indicates corresponding diseases (if any) from the current ECG. Table 2 shows the normal values of ECG and Table 3 shows the abnormal ECG values with associate diseases.

Primarily, heart diseases that the proposed system can detect are enlisted below with the detection technique and Figure 5(a,b,c,d,e) shows the enlisted diseases waveform.

1) Tachycardia: Resting heart rate exceeds more than 100 bpm. But the upper limit is 150 bpm. From R-R interval the heart rate (HR) can be calculated to detect tachycardia.

2) Bradycardia: HR falls less than 60 bpm and can be detected as mentioned above.

3) Hypercalcemia: QTc interval time is less than 0.32 sec

4) Hypocalcemia: QTc interval time is greater than 0.44 sec

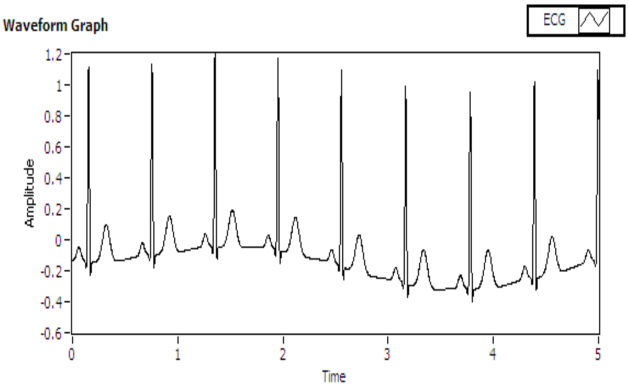
5) Atrioventricular block: PR interval is greater than 0.20 sec

TABLE II
Normal ECG parameters

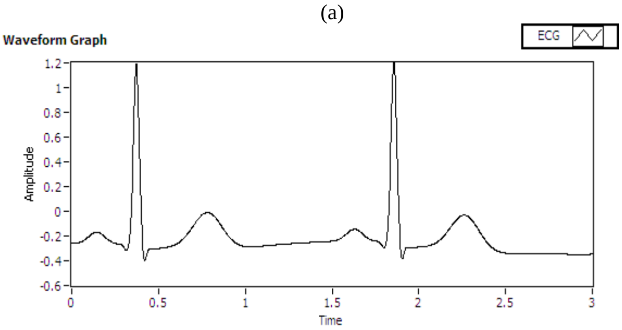
Parameter	**Duration (sec)**
P wave	0.06-0.11
PR Interval	0.12-0.20
PR Segment	0.08
QRS Complex	<0.12
ST Segment	0.12
QT Interval	0.36-0.44
T wave	0.16

TABLE III
Abnormal ECG parameters and its Effect

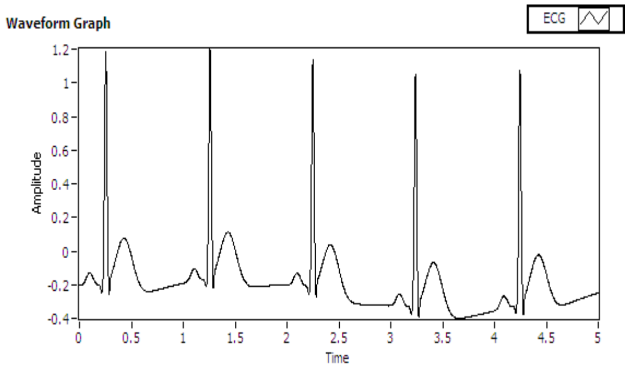
Abnormal parameter	**Heart disease**
Increased HR	Tachycardia
Decreased HR	Bradycardia
Increased PR	AV block
Long QT interval	Hypocalcemia
Short QT interval	Hypercalcemia



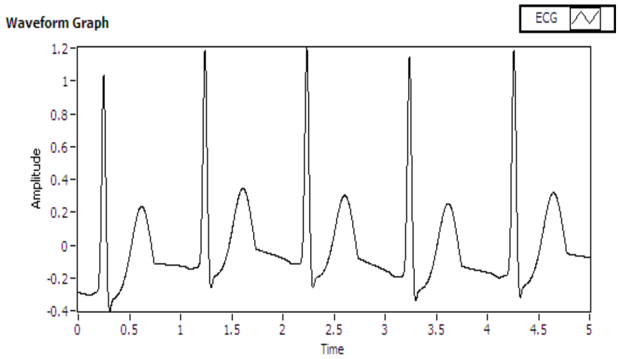
(a)



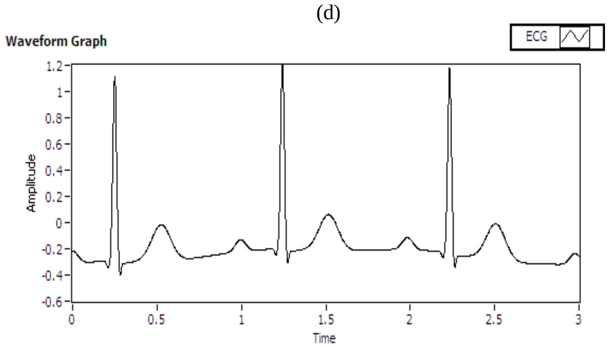
(b)



(c)



(d)



(e)

Fig 5. a) Tachycardia b) Bradycardia c) Hypercalcemia d) Hypocalcemia e) AV block

IV. Test Results

When tested on both MIT-BIH database and synthetic ECG, the proposed system successfully extracted ECG parameters and detected above mentioned heart diseases which reflect a positive sign. The remote transmission module also tested successfully to ensure proper data transmission and real time doctor-patient communication. Figure 6 shows the remote monitoring system.

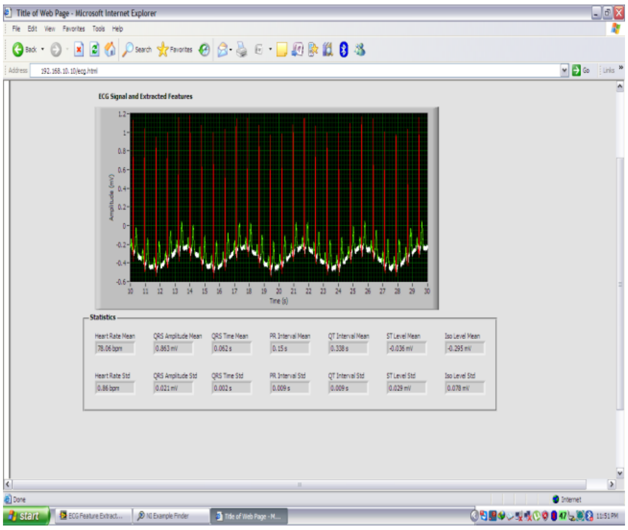


Fig 6. Remote monitoring panel on the web browser

V. Conclusion

An effective and fast automatic ECG signal analyzing system has been presented in this paper. The system is flexible and the number of technique to detect diseases can be increased without any complexity. The proposed system can be very useful for an early detection of heart abnormalities which can help a cardiovascular patient from suffering. Thus, this system can open opportunity to provide medical services in the rural area where no medical healthcare center is found in the vicinity

Acknowledgement

The author would like to thank Independent University,Bangladesh (IUB) for the fund and also Dr. Mustafa Habib Chowdhury for providing different ECG related study materials that greatly assisted to design the proposed system.

References

[1] B.U. Kohler, e. Henning, R. Orglmeister, The principles of software QRS detection, IEEE Eng. Med. BioI. 21 (I) (2002) pp. 42-57. B., Bohus C.A., Crowl, L.A. & Shor, M.H. "Distance Learning Applied to Control Engineering Laboratories". IEEE Transactions on Education, 39, 3, pp320-326 ,1996

[2] R. Silipo, C. Marchesi, Artificial neural networks for automatic ECG analysis, IEEE Trans. Signal Process. 46 (5) (1998) pp. 1417-1425.

[3] R.M. Rangayyan, Biomedical Signal Analysis: A Case-study Approach, Wiley-Interscience, New York, 2001, pp. 18-28.

[4] D. Benitez, P.A. Gaydecki, A Zaidi, A P. Fitzpatrick, The use of the Hilbert transform in ECG signal analysis, Comput. Biol.Med. 31 (2001) pp. 399-406.

[5] 1. Pan, W.J. Tompkins, A real-time QRS detection algorithm, IEEE Trans. Biomed. Eng. BME-32 (3) (1985) pp. 230-236.

[6] K.V. Suarez, J.e. Silva, Y. Berthoumieu, P. Gomis, M. Najim, ECG beat detection using a geometrical matching approach, IEEE Trans. Biomed. Eng. 54 (4) (2007) pp. 641-650

[7] X. Xu, Y. Liu, ECG QRS complex detection using slope vector wavefonn (SVW) algorithm, in: Proceedings of the 26tl' International Conference of the IEEE EMBS, 2004, pp. 3597-3600.

[8] Islam, M. R., Ahmad, S. , Hirose, K., & Molla,M. K. 1. (May 30 2010- June 2 2010). Data adaptive analysis of ECG signals for cardiovascular disease diagnosis. Paper presented at the Circuits and Systems(ISCAS), Proceedings of 2010 IEEE International Symposium.

[9] Chr. Zywietz. "A Brief History of Electrocardiography -Progress through Technology" Biosigna Institure for Biosignal Processing and Systems Research, Hannover, Germany.

[10] Carlos.C "Heart Rate Monitor and Electrocardiograph Fundamentals", Free SCale Semiconductor Application Note, Documen Number: AN4059, Rev.O, 3/2010.

[11] Bruce Carter, "Filter Design in Thirty Seconds", Texas Instruments Application Report SLOA093 December 2001.

[12] Himanshu. S, 1.Sam. 1. K, Y.Ashok and Dr.A.Vimala Juliet, "Advanced ECG signal Processing using Virtual Instrument", International J. of Recent Trends in Engineering and Technology, Vol.3, No.2, May 2010, pp111-114.

[13] Snellen, H.A. Willem Einthoven (1860-1927) - Father of Electrocardiography. Life and Work, Ancestors and Contemporaries, Kluwer Academic Publishers, Dordrecht/ Boston/ London 1995.

[14] Mark E. Silverman, M.D. Will em Einthoven - The Father of Electrocardiography. Emory University School of Medicine and Piedmont Hospital, Atlanta, Georgia, USA

[15] S. Pavlopoulos, R. H. Istepanian, S. Laxminarayan, and C. S. Pattichis, “Emergency health care systems and services: section overview,” in Proceedings of the 5th International IEEE EMBS Special Topic Conference on Information Technology Applications in Biomedicine, pp. 371–374, 2006.

[16] R. Anta, S. El-Wahab, and A. Giufrida, Mobile Health: The potential of Mobile Telephony to Bring Health Care to the Majority, Inter-American Development Bank, 2009

[17] F. Touati, R. Tabish, and A. B. Mnaouer, “A real-time BLE enabled ECG system for remote monitoring,” APCBEE Procedia, vol. 7, no. 0, pp. 124 – 131, 2013, the 3rd International Conference on Biomedical Engineering and Technology - ICBET 2013

[18] ' S.Correia, 1. Miranda, L.Silva, and ABarreto, "Labview and MATLAB for ECG Acquisition, Filtering and Processing", 3'd International Conference on Integrity, Reliability and Failure, Portol 24 July 2009.

[19] M. Abo-Zahhad, SabahM. Ahmed, and O. Elnahas “A Wireless Emergency Telemedicine System for Patients Monitoring and Diagnosis” Hindawi Publishing Corporation International Journal of Telemedicine and Applications Volume 2014, Article ID 380787, 11 pages

[20] Mr. Bhavin Mehta, Ms.Divya Rengarajan, Mr. Ankit Prasad” Real Time Patient Tele-monitoring System Using LabVIEW” International Journal of Scientific & Engineering Research, Volume 3, Issue 4, April-2012

[21] Mitrra Potheri Ramesh, Shruthi Balasubramanian, Varsha Vijayan, Geethanjali Balasubramanian, Mahesh Veezhinathan “Design and Development of a Two Channel Telemedicine System for Rural Healthcare” Scientific Research Publishing (SCIRP) 2013, 5, 579-583

[22] Nishigandha D. Agham, Vijaya R. Thool, Ravindra C. Thool “Mobile and Web Based Monitoring of Patient’s Physiological Parameters using LabVIEW” 2014 Annual IEEE India Conference

[23] National instruments Corporation. LabVIEW User Mannual.

[24] National instrument LabVIEW ecg signal processing.

[25] Channappa Bhyri, Kalpana. V, S.T.Hamde, and L.M. Waghmare "Estimation of ECG features using Lab VIEW" TECHNIA - International Journal of Computing Science and Communication Technologies, VOL. 2, NO. I, July 2009. (ISSN 0974-3375)

[26] Hong min Wang , Dan dan Li, Ping Xue, Jie Zhu and Hai bo Li "LabVIEW-based data acquisition system design" 2012 International Conference on Measurement, Information and Control (MIC)

[27] Ying-Chien Wei, Yu-Hao Lee, and Ming-Shing Young "A Portable ECG Signal Monitor and Analyzer"

[28] Chouhan Y.S. and Mehta S.S., "Threshold-based Detection of P and T- wave in ECG using New Feature Signal", International Journal of Computer Science and NetworkSecurity, Vo1.8, No.2, February 2008.

[29] www.physionet.org. accessed on March 2016

GPON Triple Play and SDH Connectivity Structure with Cost Analysis

Md. Hayder Ali
Institute of Information and Communication Technology (IICT)
Bangladesh University of Engineering and Technology (BUET)
hayder.iict@gmail.com

Md. Saiful Islam
Institute of Information and Communication Technology (IICT)
Bangladesh University of Engineering and Technology (BUET)
mdsaifulislam@iict.buet.ac.bd

***Abstract*— The rapid growth of bandwidth requirements for services like IP television and video on demand over Internet together with high speed Internet access have demand for very high bandwidth to customers as well as the changing role of enterprise networking are causing disruptive change in the enterprise local area networks. The most suitable solution for satisfying the high bandwidth demand with a long reach is using optical cable to customers through gigabit passive optical network (GPON) technology. In the last one decade many research work has been carried out on network architecture, transmission mechanisms, power budget, bandwidth allocation and scalability of GPON technology. But to tape the full potential of GPON and extends its last mile limit, there is no detail analysis regarding the convergence of synchronous digital hierarchy (SDH) connectivity as well as which particular wavelength should be optimum for transmission. In this paper a new enhanced GPON architecture is proposed incorporating SDH transmission at optimum wavelength.**

***Keywords -FTTH (Fiber to the Home), LDP (Local Distribution Point), Optical Network Terminal (ONT), ODN (Optical Distribution Network), OLT Erbium-doped fiber amplifiers (Optical Line Terminal), BDB (Building Distribution Box),* EDFA(Erbium-Doped Fiber Amplifiers), SDH (Synchronous Digital Hierarchy).**

I. Introduction

Data growth in telecom market has reduced the prominence of traditional wire line broadband technologies such as digital subscriber line and cable modem. These technologies are not efficient enough to meet the customers' demand for high-bandwidth applications such as high speed internet access, video-on-demand, high definition TV, IPTV and online gaming. In this scenario, fiber-to-the-home (FTTH) by GPON technology, which offers advantages like high bandwidth capacity and the delivery of high speed, high quality and multi-play services (data, voice and video) through a single channel, presents a strong business opportunity for telecom operators. Full Service Support, including voice (TDM), Ethernet, ATM, leased lines, and others. Strong Operations, Administration, Maintenance, and Provisioning (OAM&P) capabilities offering end-to-end service management. The GPON technology was developed to provide high speed Ethernet services for residential and small business customers. It increases the access layer bandwidth and builds a sustainable-development access layer network. OAN (Optical Access Network) adopts technologies: active point-to-point

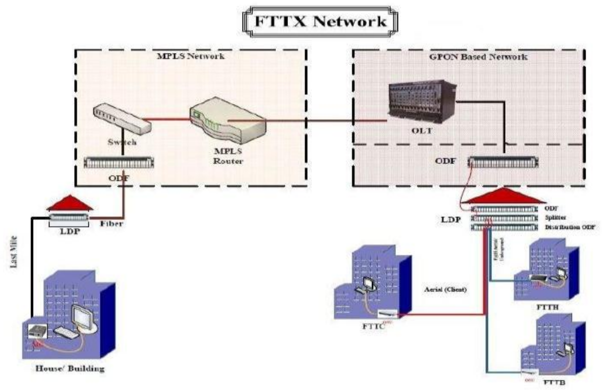


Figure 1: Open Access Network Structure (FTTx)

(P2P) Ethernet and passive optical network (PON). There are many common subsets of FTTx like- FTTN (fiber to the node or fiber to the neighbourhood), FTTC (fiber to the curb or fiber to the cabinet), FTTP (fiber to the premises), FTTB (fiber to the building or fiber to the basement), FTTH (fiber to the home) etc.

The above figure shows that if Splitted fiber directly goes to client ode/Premises/ Home then client will enjoy the device dedicatedly and if splitted fiber goes to Building's basement then from ONU/ONT client will enjoy their connectivity by short UTP cable.

The rest of this paper is organized as follows. The background studies is introduced in Section 2. In Section 3, the cost calculation is compared between Ethernet connectivity and GPON. In Section 4, modified triple play architecture is briefed. Simulation and Performance analysis are shown at section 5 and 6. Finally Section 7 draws a conclusion to this paper.

II. Background Studies

G.984.x Recommendations provide a typical GPON system model, which consists of optical line terminal (OLT), optical distribution network (ODN) and optical network unit (ONU)/ Optical network terminal (ONT). OLT is responsible for ONU/ONT upstream bandwidth allocation, and it is a central issue to allocate the bandwidth more reasonable [1]. I.Cale, A. Salihovic, M. Ivekovic [2] explained overview of Gigabit PON and analyses network architecture, transmission mechanisms and power budget in GPON systems. But there is no idea regarding SDH connectivity. M. Leo, M. Trotta [3] mentioned an alternative solution based on Wavelength Division Multiplexing (WDM-PON) that seems to have more performances than GPON in terms of bandwidth allocation, scalability and capability of unbundling. There is no idea or on how wavelength should be optimum. Ricciardi, S.; Santos-Boada, G.; Careglio, D.; Domingo-Pascual, J [4] shown an analysis between Ethernet Point to Point (EP2P) and GPON connectivity. It just an analysis of GPON general architecture. It didn't mentioned about complex architecture (like-SDH, data, voice and video connectivity) of GPON. J. Lee, I. Hwang, A. A.Nikoukar, and A. T.Liem [5] mentioned only for bandwidth allocation scheme for general triple play architecture, there should a scope for discussing SDH bandwidth allocation. S. Milanovic [6] explored an opportunity to adopt Passive Optical LANs (POLs), based on Gigabit Passive Optical Network technology (GPON), rather than continuing with use of traditional two- or three-tier switched Ethernet solution. Mostly focused on Passive optical LANs. H. Nusantara, F. Dairianta [7] analyzed the design of fiber access network systems using GEPON technology for HRB. GPON for HRB are designed to comply both for power budget and rise time budget standard. Mostly discussed with Splitting ratio for GPON system. E. J. C. González; M. E. Morocho Cayamcela [8] analyzed the integration of HSI, VoIP and IPTV services into the optical network owned by the National Telecommunications Corporation of Ecuador. It described a little bit regarding convergence of technology but it did not discussed about SDH network over GPON. M. Irfan, M. S. Qureshi, S. Zafar [9] explained an evaluation is performed of 2.5Gbps bi directional GPON based Fiber-To-The-Home (FTTH) link using advanced modulation formats. Mostly described on mobile back haul network and a single wavelength and two wavelengths were used for triple-play services with different modulation schemes. A. Vesco, R. M. Scopigno, E. Masala [10] illustrated the advantages of Time-Division Unbalanced Carrier Sense Multiple Access (TDuCSMA) in such a scenario compared to the Enhanced Distributed Channel Access (EDCA), currently provided by the IEEE 802.11 standard, in terms of both performance from the end user's point of view and network resource utilization. There was an scope for discussing about SDH transmission system over GPON system. J. Frnda, M. Voznak, P. Fazio, J. Rozhon [11] worked with Network performance simulation and quality of triple play service in IP networks. Mostly worked with queuing policy and transmission speed of interface on routers on packet networks. T. Rokkas [12] explained the cost for the deployment of a PON FTTH network is calculated in terms of NPV, IRR and payback period. A comparison is made between three PON technologies: GPON, XGPON and NG-PON2. Different scenarios regarding population density and bitrates are examined. S. S. W. Lee; K. Y. Li; M. S. Wu [13] implemented all the OpenFlow functions including- packet forwarding, bandwidth metering, statistical data collection, and status reporting. The experimental results show that the GPON virtual switch can correctly perform all the functions defined in the OpenFlow 1.3 specification.

There was an opportunity to work with SDH over GPON Network. But there are no clear idea about that. SDH connection through an optimum wavelength is important for GPON connectivity which will enhance the GPON triple Play architecture and as well as performance will also increase.

III. GPON SYSTEM

GPON or Gigabit Passive Optical Network is an optical technology based on the industry standard ITU-TG.984x which was ratified in 2003. This technology was originally developed to provide high speed Ethernet services for residential and small business customers. It supports higher rates, enhanced security, and choice of Layer 2 protocol (ATM, GEM, and Ethernet). A passive optical network (PON) is a point-to-multipoint, fiber to the premises network architecture in which unpowered optical splitters are used to enable a single optical fiber to serve multiple premises, typically 16-128. A PON consists of an optical line terminal (OLT) at the service provider's central office and a number of optical network units (ONTs, ONUs) near end users. A PON reduces the amount of fiber and central office equipment required compared with point-to-point architectures. A passive optical network is a form of fiber-optic access network.

GPON has a downstream capacity of 2.488 Gb/s and an upstream capacity of 1.244 Gbp/s that is shared among users. Encryption is used to keep each user's data secured and private from other users. Although there are other technologies that could provide fiber to the home, passive optical networks (PONs) like GPON are generally considered the strongest candidate for widespread deployments. It provides unprecedented bandwidth (shared by up to 128 premises), and a greater distance from a central office (20 to 40 kilometers), allowing service providers to enable bandwidth-intensive applications and establish a long-term strategic position in the broadband market. In downstream GPON do broadcast to all of Connected ONU/ONT.

Enterprise GPON is also a carrier class technology that provides a high level of Quality of Service (QOS) 99.999% for those customers with mission-critical requirements. GPON manufactures are now working on devises that will allow up to 10Gbs on bandwidth. In Upstream GPON use TDMA.

As a result, the a new standard known as G987 or also known at 10-PON has 10 Gbit/s downstream and 2.5 Gbit/s upstream – framing is "G-PON like" and designed to coexist with GPON devices on the same network. This is great news for data network managers looking for low-cost, high-bandwidth, networking technologies in order to keep up with the demands on data applications and growth including "cloud" services. By GPON Technology service provider could provide several service to its customers like- IP TV, Voice (VoIP), Video, Data Connectivity, Internet connectivity, value added service (Online gaming, Social networking, Video on Demand etc) and other services.

IV. COST CALCULATION

Each switch increases the carbon footprint of the organization. The lesser the efficiency of the switch, the greater the footprint. According to PG&E, 0.524 pounds (lb) of carbon dioxide (CO_2) are emitted for every kWh of power consumed. A 100W switch running 24 hours a day emits close to 569 lb of CO_2 every year. Such emissions increase the carbon footprint of an organization drastically. Thus, there exists a strong business and environmental need to study the power consumption of Ethernet switches. However an approximate cost calculation is given bellow for Ethernet Connectivity.

TABLE 1: ETHERNET COST CALCULATION

Ethernet Connectivity Cost: Passive Device (Including Outside Planning)		
Particulars	**Cost (USD) (Approximate)**	**Descriptions**
Patch Cord: Switch Port to LDP	5.00	20M Patch Cord Price
ODF Cost at CO	4.00	144 Port ODF: 700 USD
Space Cost at CO	3.00	42U Open Rack: 106 USD. Considering 2U Price: 42.8. 1U for ODF and 1U for Switch
Under Ground Fiber: CO to LDP	280.00	Considering 244 Core Cable Average Distance OLT/ODF to Splitter: 4,000 Meter Underground Fiber Cost/Meter: USD 15.00 (Consisting of 216 Core Fiber, Duct & Fiber Laying Cost)
ODF Cost at LDP/Port	2.00	24 Port ODF: 70 USD Considering pigtail and adapter
LDP Space Cost/Port	2.00	1U LDP Space: 20 USD
ODF Installation Cost/Port	1.00	1U ODF Installation: 10 USD
Per Connectivity Cost	**297.00**	

The protection for PON is very important to increase reliability. Meanwhile, access network providers need to keep capital and operational expenditures (CAPEX and OPEX) low in order to be able to offer economical solutions for the customers. Thus, minimizing the cost for network protection while maintaining an acceptable level of connection availability is an important challenge for the current fiber access networks. An approximate cost calculation for GPON connectivity is given bellow-

TABLE 2: FTTX COST CALCULATION

FTTx per Client Cost (Including OSP)		
Particulars	**Unit Price (USD) (Approximate)**	**25 USD per Client Cost (1:32 Splitter)**
OLT Chesis including dual Power	1,400.00	
Packet Switching and CPU Management	2,000.00	
8-port GPON ports with SFP type line card	7,500.00	**20 USD per Client Cost (1:64 Splitter)**
2 Port Gigabit Ethernet Unit	70.00	
PON SFP Module	300.00	
GE Uplink SFP Module	100.00	

FTTH (Fiber To The Home) connectivity scenario-

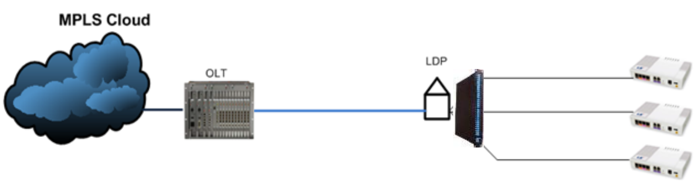


Figure 2: FTTH Connectivity Structure

TABLE 3: FTTH COST CALCULATION

Access Network	**Central Office**	**OSP – Fiber Optic**	**CPE**	**Total Cost**
GPON	**70.00 % (Less)**	**92.00 % (Less)**	**130.50 % (Higher)**	**49.00 % (Less)**

FTTB (Fiber To The Building) connectivity scenario-

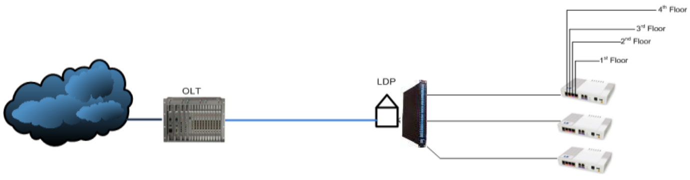


Figure 3: FTTB Connectivity Structure

TABLE 4: COST CALCULATION FOR FTTB CONNECTIVITY

Access Network	**Central Office**	**OSP – Fiber Optic**	**CPE**	**Total Cost**
GPON	93.50 % (Less)	98.95 % (Less)	60 % (Less)	**88.50 % (Less)**

V. Modified Triple Play Architecture

The triple-play service is realized as a combination of data, voice, and video signals. The high-speed internet component is represented by a data link with 1.25 Gb/s downstream bandwidth. A traditional triple play architecture is like as bellow-

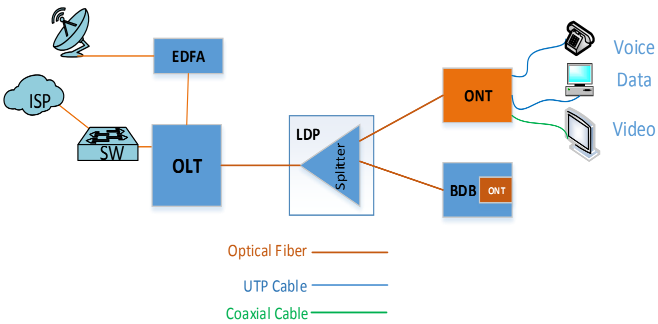


Figure 4: Traditional Triple Play Architecture

Instead of using EDFA combiner we could use MUX and direct modulated laser and could get the best performance for optimum wavelength.

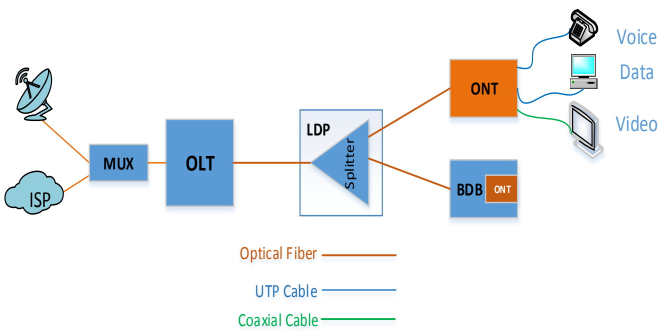


Figure 5: Enhanced Triple Play Architecture

VI. SIMULATION

We have done the simulation by using OptSIM Simulation software. The simulation architecture has mainly two part, The OLT block and The ONT block.

OLT block (Transmitter block) consists of Data/VOIP and Video components. The Data/VOIP transmitter modeled with pseudo-random data generator (PRBS), NRZ modulator driver, direct-modulated laser, and booster amplifier. The video component modeled as RF SCM (sub-currier multiplexed) link with only two tones (channels) for simplicity. The two channels we used are from standard NTSC analog CATV frequency plan - channel 2 and channel 78 at frequencies 55.25 MHz and 547.25 MHz, respectively. RF video transmitter consists of two Electrical Signal Generators, summer, direct-modulated laser, and pre-amplifier. Next, Data/Voice and Video signals are multiplexed at Multiplexer and launched into 20-km fiber span. Output from the fiber trunk goes through the 1:16 splitter and then to individual users. User’s ONT consists of splitter and data and video receivers. Data receiver configured with optical filter, PIN/TIA receiver, and BER Tester. The video signal receiver consists of optical filter, PIN/TIA receiver and electrical filters.

The ONT block (Receiver block) can be represented as VOIP service (voice over IP, packet-switched protocol) and can be combined with data component in physical layer simulations. Finally, the video component can be represented as a RF video signal (traditional CATV) or as IPTV signal that also can be combined with data. To modify the traditional triple play service, we consider the former case with RF video link. To optimize the bandwidth in PON the transmission through the optical fiber path employs the CWDM technique with data/voice component transmitted at wavelengths in the range of 1480-1500 nm, and video within the 1550-1560 nm range.

VII. PERFORMANCE ANALYSIS

In first phase all voice, data and video signals are at same frequency and store the results and at second phase different signal combinations are simulated and results are stored. In first phase simulation there were no measurable outcome as there used same wavelength both for Data + voice and Video. There were different outcome at second phase simulation.

Second phase Simulation was like bellow-

TABLE 5: SECOND PHASE SIMULATION WAVE LENGTH

Data+Voice	Video
1310	1490
1310	1550
1490	1310
1490	1550
1550	1310
1550	1490

After both phase simulation we got the best result for 20 Km distance. We got the 1490 nm wave length for data and voice and 1550 nm for video are given the best output signals.

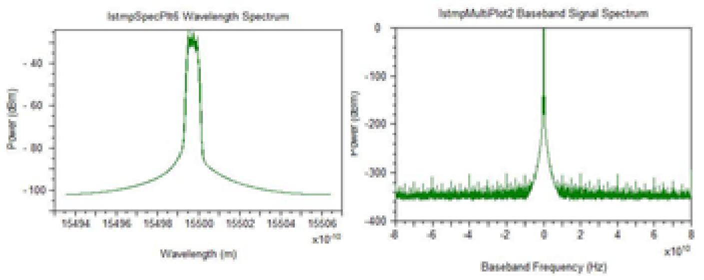


Figure 6: Output Signal_ Video and Baseband Electrical Signal after Electrical filtering

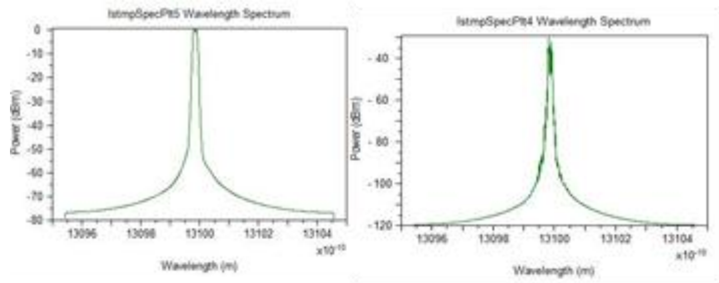


Figure 7: Input and Output Signal_ Data+Voice

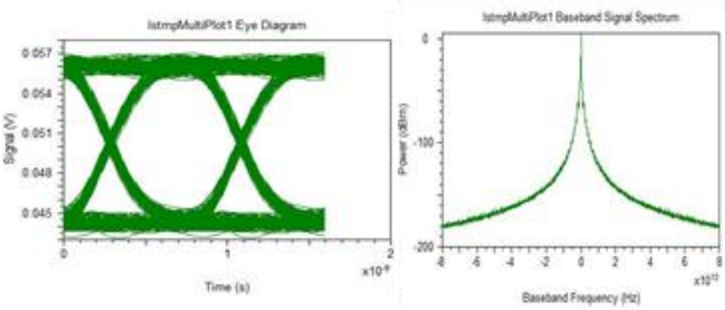


Figure 8: Output Eye Diagram and Baseband Signal_ Data+Voice

VIII. CONCLUSIONS

Analyzing various case for several wave length for Data, voice and video, it is found that by using Direct modulated laser and an optical MUX instead of signal combiner (EDFA) 1490 nm wave length for Data+Voice and 1550 nm for Video are best for long distance triple play connectivity in terms of performance. To achieve the broadband targets set by the government under the National Telecom Policy, it will be important to drive FTTH growth along with other technologies. GPON, through the Generic Framing Procedure (GFP)-based adaptation method, offers a clear migration path for adding services onto the PON without disrupting existing equipment or altering the transport layer in any way. GPON Connectivity is more efficient then Ethernet connectivity. It is possible to provide 7/8 GPON Connectivity by the cost of one Ethernet connectivity. Quad play service (SDH, voice, video and data) is possible from single device. Industry based implantation could be future work.

REFERENCES

[1] ITU-T Recommendation G.984.1, G.984.2, G.984.3, G.984.5, and G.984.6, Gigabit-capablepassive optical networks (GPON): General characteristics, Physical media dependent layer specification, Transmission convergence layer specification, Enhancement band, Reach extension (ex G.984.re-GPONoptical reach extension), 2003-2007.

[2] I. Cale, A. Salihovic, M. Ivekovic; "Gigabit Passive Optical Network – GPON", Proceedings of the 29th International Conference on Information Technology Interfaces, Croatia, 25-28 June 2007.

[3] Leo, M.; Trotta, M.," Performance evaluation of WDM-PON RSOA based solutions in NGAN scenario " , Proceedings of the 50th FITCE Congress (The Forum for European ICT & Media Professionals), Italy, 31 August- 3 September 2011.

[4] Ricciardi, S.; Santos-Boada, G.; Careglio, D.; Domingo-Pascual, J., "GPON and EP2P: A Techno-Economic Study", Proceedings of the 17th European Conference on Networks and Optical Communications (NOC), Spain,20- 22 June 2012.

[5] J. Lee, I. Hwang, A. A.Nikoukar, and A. T.Liem "Comprehensive Performance Assessment of Bipartition Upstream Bandwidth Assignment Schemes in GPON" . Journal of Optical Communications and Networking ,vol.: 5, no. 11, pp. 1285-1295, November 2013.

[6] S. Milanovic, "Case Study for a GPON Deployment in the Enterprise Environment", Journal of Networks, vol. 9, no. 1, pp-42-47, January 2014.

[7] H. Nusantara, F. Dairianta; "Design and Analysis of FTTH - GEPON for High Rise Building", Proceedings of the 8th International Conference on Telecommunication Systems Services and Applications (TSSA), Indonesia, 23-24 October 2014.

[8] E. J. C. González; M. E. Morocho Cayamcela; "Integration of a Triple-play platform service to the GPON infrastructure of the National Telecommunications Corporation of Ecuador", Proceedings of the Scientific and Technical Conference of the Andean Council of IEEE, Bolivia, 15-17 October 2014.

[9] M. Irfan, M. S. Qureshi, S. Zafar; "Evaluation of Advanced Modulation Formats using Triple-Play Services in GPON Based FTTH", Proceedings of the International Conference on Cloud Computing (ICCC),KSA, 27-28 April 2015.

[10] A. Vesco, R. M. Scopigno, E. Masala; "TDuCSMA: Efficient Support for Triple-Play Services in Wireless Home Networks", Proceedings of the IEEE International Conference on Communications (ICC), UK, 8-12 June 2015.

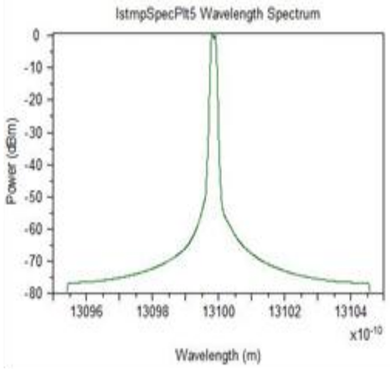
[11] J. Frnda, M. Voznak, P. Fazio, J. Rozhon; "Network Performance QoS Estimation", Proceedings of the 38th International Conference on Telecommunications and Signal Processing (TSP), Czech Republic, 9-11 July 2015.

[12] T. Rokkas; "Techno economic analysis of PON architectures for FTTH deployments", Proceedings of the Conference of Telecommunication, Media and Internet Techno-Economics (CTTE), Germany, 9-10 November 2015.

[13] S. S. W. Lee; K. Y. Li; M. S. Wu; "Design and Implementation of a GPON-based Virtual Open Flow-enabled SDN Switch", Journal of Lightwave Technology, issue: 99, pp:1, 2016. [3] H. Xie, T. Xiaodong, Z. Li, "An algorithm to implement dba of GPON", Proc. SPIE 5626, Network Architectures, Management, and Applications II, 1173,vol.2 , no.6, pp. 58-59, February 28, 2005.

Output Signal Analysis at end user side

Video_Data_Voice_1310 nm



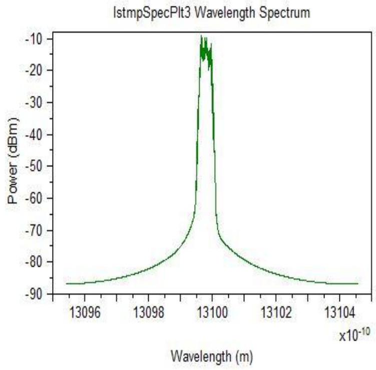
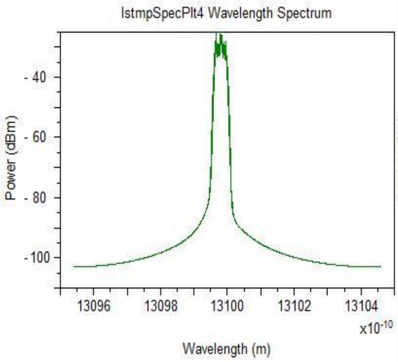
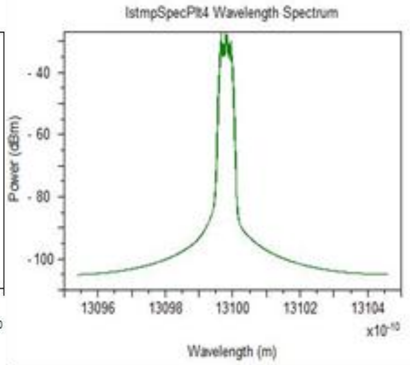


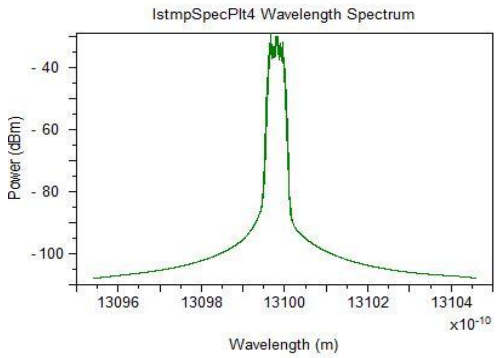
Fig. Input Signal:Data+Voice and Video (1310 nm)



After 1 KM Distance



After 10 KM Distance

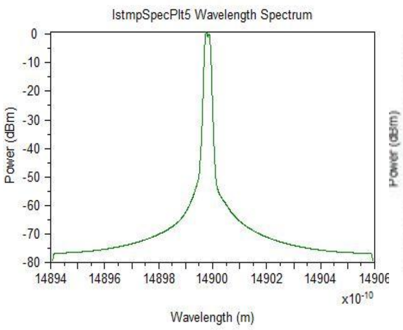


After 20 KM Distance

Fig. Output Signal: Data+Voice for different distance (1310 nm)

Output Signal for Video: There is no output signal for Video as using same frequency both for data+Voice+ Video (1310 nm).

Video_Data_Voice_1490 nm



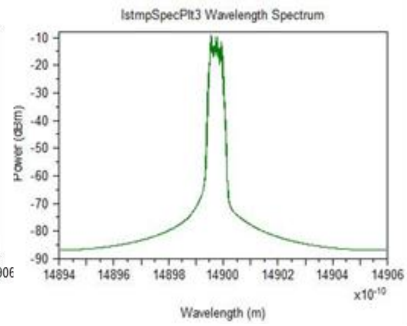
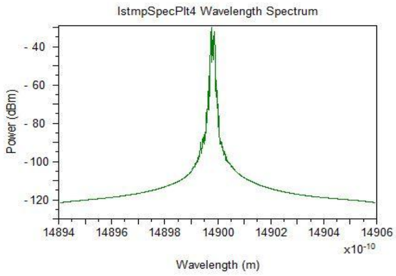
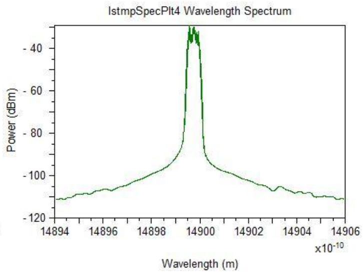


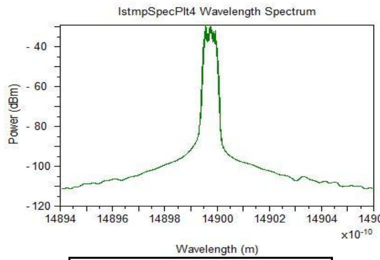
Fig. Input Signal: Data+Voice and Video (1490 nm)



After 1 KM Distance



After 10 KM Distance

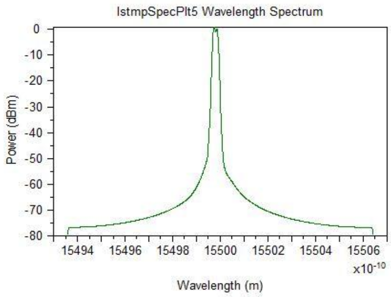


After 20 KM Distance

Fig. Output Signal: Data+Voice for different distance (1490 nm)

Output Signal for Video: There is no output signal for Video as using same frequency both for data+Voice+ Video (1490).

Video_Data_Voice_1550 nm



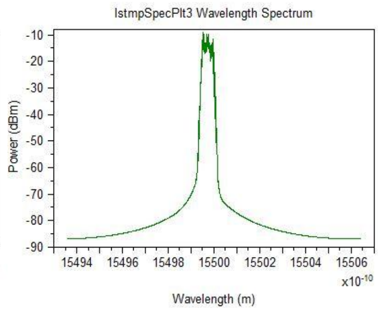
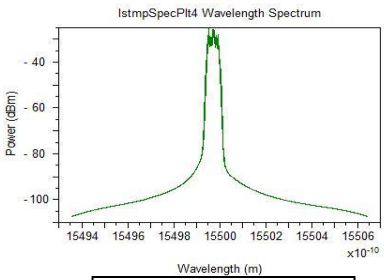
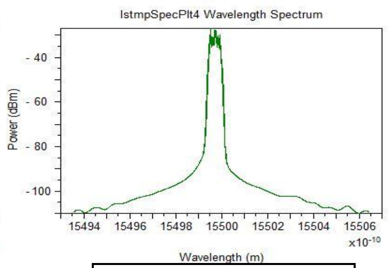


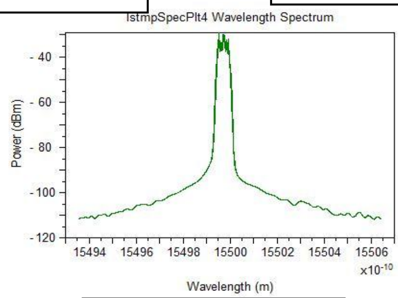
Fig. Input Signal: Data+Voice and Video (1550 nm)



After 1 KM Distance



After 10 KM Distance



After 20 KM Distance

Fig. Output Signal: Data+Voice for different distance (1550 nm)

Video (1310 nm)_Data+Voice(1490 nm)

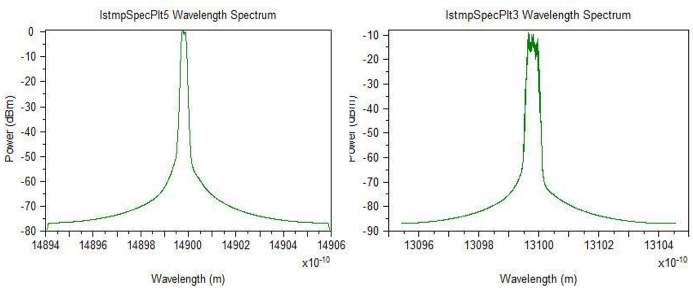
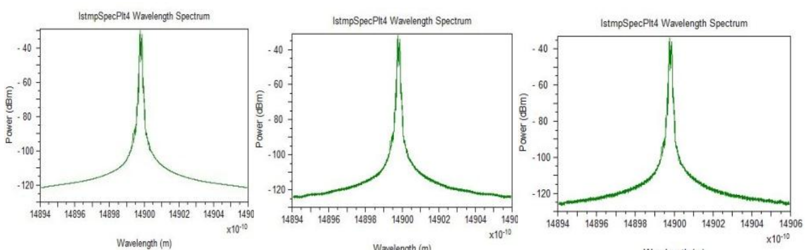
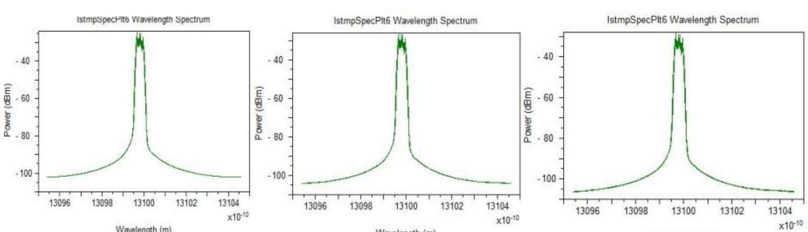


Fig. Input Signal: Data+Voice (1490 nm) and Video (1310 nm)



After 1 KM Distance | After 10 KM Distance | After 20 KM Distance

Fig. Output Signal: Data+Voice for different distance (1490 nm)



After 1 KM Distance | After 10 KM Distance | After 20 KM Distance

Fig. Output Signal: Video for different distance (1310 nm)

Video (1310 nm)_Data+Voice(1550 nm)

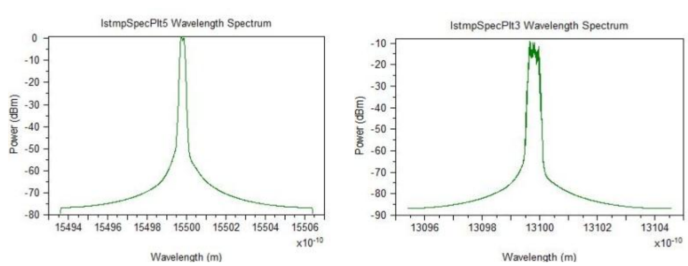
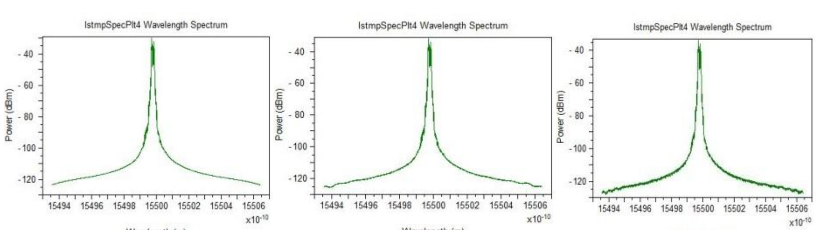
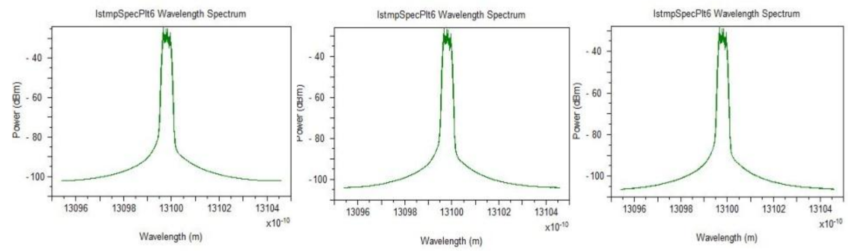


Fig. Input Signal: Data+Voice (1550 nm) and Video (1310 nm)



After 1 KM Distance | After 10 KM Distance | After 20 KM Distance

Fig. Output Signal: Data+Voice for different distance (1550 nm)



After 1 KM Distance | After 10 KM Distance | After 20 KM Distance

Fig. Output Signal: Video for different distance (1310 nm)

Video (1550 nm)_Data+Voice(1490 nm)

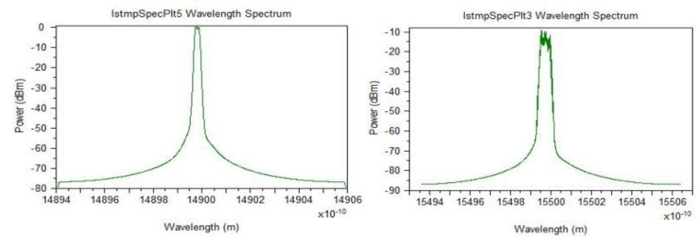
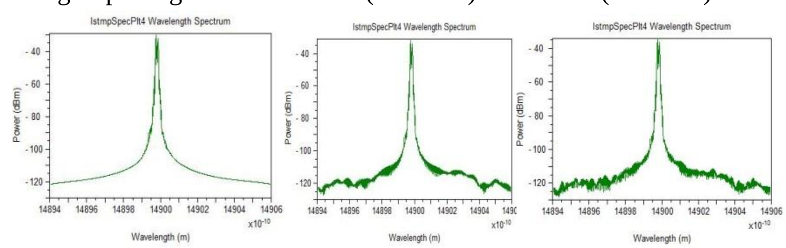


Fig. Input Signal: Data+Voice (1490 nm) and Video (1550 nm)



After 1 KM Distance | After 10 KM Distance | After 20 KM Distance

Fig. Output Signal: Data+Voice for different distance (1490 nm)

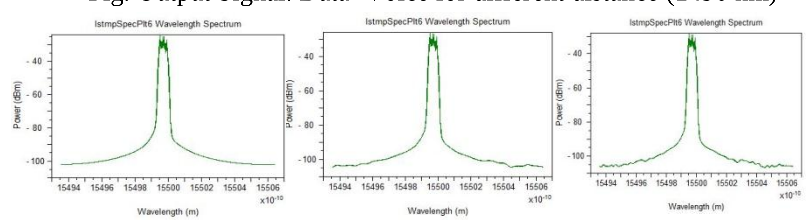


Fig. Output Signal: Video for different distance (1550 nm)

Eye Diagram Comparison

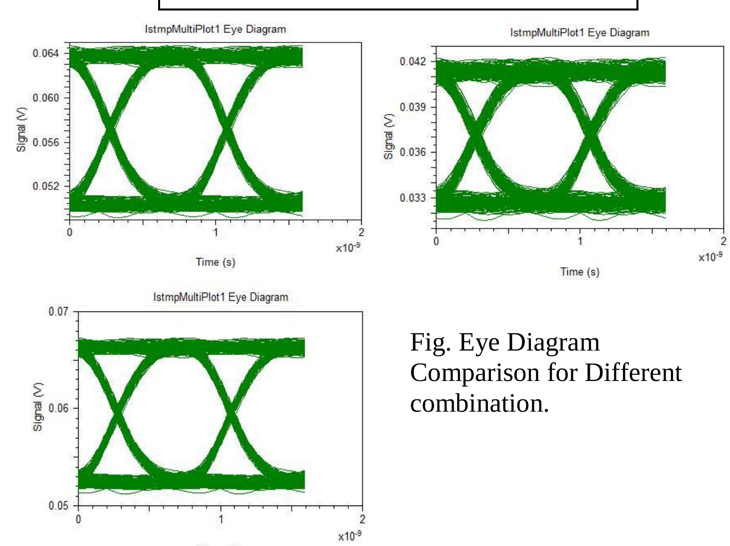


Fig. Eye Diagram Comparison for Different combination.

Basic Sequential Algorithmic Scheme Based Blind Common Phase Error Compensation in OFDM Systems

Md. Alamgir Hossain
Khulna University of Engineering & Technology
Khulna-9203, Bangladesh
Email: mah@mail.kuet.ac.bd

***Abstract*—In this paper, Basic Sequential Algorithmic Scheme (BSAS) based blind Common Phase Error (CPE) compensation method is proposed to abate the effect of Phase Noise (PHN) in OFDM system to retain the system bandwidth. By using the proposed algorithm, all the signal points are grouped according to the constellation size and average angle of each group is estimated. The estimated average angles are compared with the corresponding ideal angles and averaged again to estimate CPE. The performance of the proposed method is demonstrated by several computer simulations using MATLAB.**

Keywords: BSAS, OFDM, Phase Noise, Common Phase Error, Inter-carrier Interference, Phase Locked Loop.

I. Introduction

OFDM system has been widely implemented in various wireless and wireline communication systems, such as IEEE 802.11a/g, IEEE 802.16, HIPERLAN etc. It is vigorous against inter-carrier interference (ICI) and inter-symbol interference (ISI) caused by multipath frequency selective channel. But, the most vital aspect, the orthogonality among the sub-carriers, of OFDM system is intimidated by the presence of PHN. To compensate for the effect of PHN, several pilot based PHN estimation algorithms are proposed, such as in [1]-[3]. In most of the existing pilot based methods, some pilots are sent with each OFDM block to estimate the multi-path channel and these pilot are also used to estimate the phase noise. The exploitation of pilot symbols reduces the system bandwidth. Besides, due to the time-varying nature of PHN, the pilot sequence requires to be transmitted periodically, results further reduction of system throughput. In order to improve the bandwidth efficiency, blind methods for PHN compensation have attracted much attention. In [4]-[13] the authors have proposed a blind algorithm that suffers from higher computational complexity. In most of the existing blind algorithms, there is a performance losses arising from decision errors, since it uses the decision-directed approach [7]. Recent advancement in wireless communication, such as WiMAX has been adopted with one out of three consecutive OFDM symbols are allocated for pilot while in LTE one full OFDM is dedicated for channel estimation once every seven time slots. Therefore, it is crucial to detect data symbols with the help of pilot symbols, this motivation directs this research to develop an algorithm to estimate the data symbol as well as phase noise without pilot in each OFDM symbol. In this paper, BSAS based [8] blind PHN compensation method is proposed to abate the effect of CPE in OFDM system caused by PHN. In the proposed algorithm, the signal points are grouped according to the constellation size by exploiting the BSAS algorithm.

II. System Model

A block diagram of OFDM transceiver system with PHN is shown in Fig. (1). The time-domain signal $s[n]$ is obtained by applying N-point Inverse Fast Fourier Transform (IFFT) on frequency-domain signals, $S[k]$ given by,

$$s[n] = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} S[k] e^{j2\pi nk/N} \tag{1}$$

where, $k, n = 0, 1, \cdots N-1$. a cyclic prefix (CP) of length $N_c \geq L$ samples are added at the beginning of each OFDM block to prevent the effect of ISI, where L is the length of the multi-path channel. The carrier frequency offset and channel conditions are estimated in the training phase and compensated for in the data detection stage, hence the effect of frequency offset is minimized and the channel state information is assumed available at the receiver. After removing the CP, the received complex baseband signal of m^{th} OFDM block in the presence of PHN can be written as

$$y_m[n] = e^{j\phi[n]} \sum_{i=0}^{L-1} h_m[i] s_m[n]_{modN} + \nu_m[n], \tag{2}$$

where, h is L taps baseband channel and ν is AWGN noise. To demodulate the received time-domain signal, N-point FFT is applied on it and hence the frequency-domain signal $Y_m[n]$

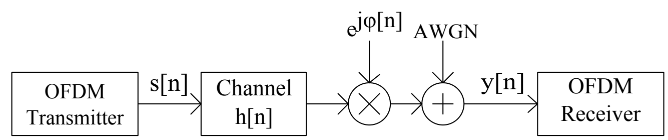


Fig. 1. OFDM Transceiver System with PHN

is given by

$$\begin{aligned} Y_m[n] &= \sum_{l=0}^{N-1} H_m[l]S_m[l]C[l-n]_N + Z_m[n] \\ &= \Phi_m[0]S_m[n]H_m[n] \\ &\quad + \sum_{l=0,l\neq n}^{N-1} H_m[l]S_m[l]C[l-n]_N + Z_m[n] \end{aligned} \quad (3)$$

It will discussed in the following sub-section that the PHN angle is small and can be expressed as $e^{j\phi[n]} \approx 1 + j\phi[n]$. Therefore,

$$\Phi_m[0] \approx \frac{1}{N}\sum_{n=0}^{N-1}(1+j\phi[n]) = 1+\frac{j}{N}\sum_{n=0}^{N-1}\phi[n] = 1+j\bar{\phi} \quad (4)$$

where,

$$\bar{\phi} = \frac{1}{N}\sum_{n=0}^{N-1}\phi[n]$$

is the angle of rotation caused by CPE. $\Phi_m[0]$ in equation (3) depict the effect of CPE that causes the rotation of every received symbol by an average angle $\bar{\phi}$ given by equation (4). The second term of equation (3) is ICI. It will be discussed later that the variation of generated PHN process is low and the effect of ICI is small compared to CPE. Besides, the ICI term can be approximated as zero-mean complex Gaussian noise, and can be merged with AWGN noise. So, the equation (3) can be rewritten as

$$Y_k = \Phi_0 S_k H_k + \xi_k \quad (5)$$

where, ξ_k consists of ICI and AWGN noise. The main objective of this research is to mitigate the effect of CPE, i,e to estimate and compensate for the average angle that causes CPE to demodulate the received data bit efficiently.

III. Phase Noise Model

A. Source of PHN

The output voltage of an ideal LO can be described mathematically as a pure sine wave of constant frequency and amplitude. In practice, this type of oscillator is impossible. A real oscillator output voltage is expressed in [9] given by

$$v(t) = [A + \alpha(t)]cos\{2\pi f_c t + \phi(t)\} \quad (6)$$

where, $\alpha(t)$ and $\phi(t)$ are amplitude and phase noise modulation respectively. If the oscillator is well-designed, amplitude noise is less significant than phase noise. Practically, it is very difficult to implement this type of oscillators. The rapidly increasing demand of bandwidth leads to higher order and more compact carrier frequency overlapping modulation techniques. It gets pretty challenging to recover the transmitted data in the presence of phase noise. Since, the oscillator amplitude noise can be fixed-up by using an automatic gain controller (AGC), the main issue in this paper is to deal with the time-varying random PHN process.

The nature of PHN rely on the type of frequency generator that may be a free running oscillator or frequency synthesizer.

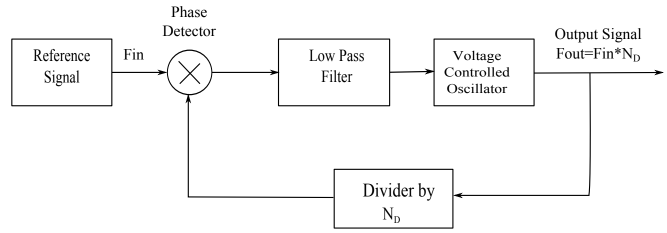


Fig. 2. Block Diagram of PLL.

Most of the wireless communication systems use phase lock loop (PLL) based frequency synthesizer [10] due to its high level of stability, easy control, wide range of frequency generation as well as higher accuracy. PHN is generated from the non-ideal characteristics of PLL results PHN that is random process caused by the phase fluctuation of the oscillators.

Better hardware design may reduce the effect of PHN but improved estimation scheme is an important issue to estimate the PHN. Most of the existing research works assume Wiener model (Random walk) for phase noise [11]. Such a model is only suggested when the carrier signal is generated from a stand alone local oscillator. This type of design is almost never adopted as a free running oscillator that slowly drift away from the required phase and frequency. Whereas, the gain of a PLL is almost constant within the loop bandwidth and there is a 30dB/dec roll off outside the loop bandwidth. The block diagram of a PLL is shown in Fig. (2) that consists of a reference signal source, a phase detector (PD), a low pass filter (LPF), a voltage controlled oscillator (VCO) and a frequency divider. The function of PD is to detect the phases of two input signals, then compare the phases and gives an error signal depending on the phase difference. If the reference input and the signal from divider are given by

$$s_r = A_r sin\{\omega_c t + \phi_r(t)\} \quad (7)$$

$$s_d = A_d cos\{\omega_c t + \phi_d(t)\} \quad (8)$$

then, the PD output will be

$$\begin{aligned} s_p = s_r \times s_d = \frac{A_r A_d}{2}&[sin\{\phi_r(t) - \phi_d(t)\} \\ &+ sin\{2\omega_c t + \phi_r(t) + \phi_d(t)\}] \end{aligned} \quad (9)$$

The high frequency component is filtered out by the LPF and the input voltage of the VCO is given by

$$s_{LP} = \frac{A_r A_d}{2} sin\{\phi_r(t) - \phi_d(t)\} \quad (10)$$

where, ϕ_r and ϕ_d are the phase of reference and detector signals respectively. Using PLL and VCO, random phase noise is generated as shown in Fig. (3) that illustrates the PHN is a zero mean random variable.

B. Effect of PHN on OFDM

To validate the consequence of PHN on OFDM systems, computer simulations were carried out using the following parameters shown in section V. It is clearly seen from the scatter plot shown in Fig. (4) that the received signals, blue

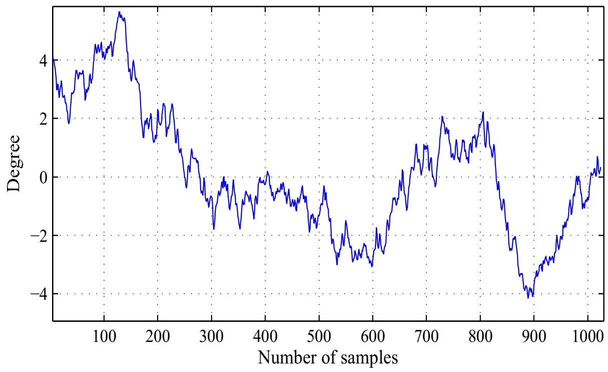


Fig. 3. The Random PHN.

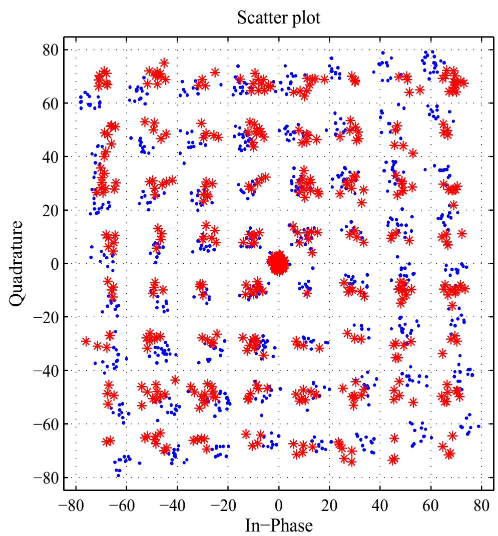


Fig. 4. Effect of CPE on 64-QAM OFDM

marked (dot), are rotated counter-clockwise from its original location, the place of signal without phase noise marked in red (star), by an average angle $\bar{\phi}$. The main objective of this research is estimate this average angle of rotation and to mitigate the effect of CPE.

IV. CPE Estimation

The signal points are rotated due to CPE and if the angle of rotation exceeds threshold angle, then it is quite difficult to detect the transmitted data symbols. Different threshold angles depending on different M-QAM constellations are discussed in [9]. In that case, different algorithms or special error correcting codes are necessary to efficiently detect the symbol. In this section, a CPE compensation algorithm is proposed to subside the effect of CPE in OFDM. In the proposed algorithm, the signal points are grouped according to the constellation size by exploiting BSAS algorithm. Then, the angles of signals of same group are calculated and compared with the idle angle and averaged to obtain phase error for that particular point. This process is repeated for all available constellation points and averaged again to estimate the angle of rotation caused by CPE. The performance of the proposed method is demonstrated by the computer simulations. PLL frequency synthesizer is used as the source of PHN [12], and for the simplicity, only the receiver phase noise is considered. The performance of the proposed algorithm is demonstrated by the computer simulations. To do this, 64-QAM modulation technique is exploited to modulate the frequency domain signal. Firstly, the conventional pilot based method of CPE estimation will be analyzed, and then the simple but robust BSAS based CPE estimation algorithm will be proposed.

A. Conventional Pilot Based Estimation

Suppose S_p presents the set of indices corresponding to pilots in an OFDM signal. To estimate the CPE angle $\bar{\phi}$, the least-square (LS) method [3] is applied by minimizing the cost function,

$$\min_{C_0} \sum_{k \in S_p} \mid Y_k - C_0 S_k H_k \mid^2$$

The pilot based estimation of CPE angle can be obtained [13] as follows-

$$\hat{\bar{\phi}}_{pilot} = \Im \left(\frac{\sum_{k \in S_p} Y_k (H_k S_k)^*}{\sum_{k \in X_p} \mid H_k S_k \mid^2} \right) \tag{11}$$

B. Proposed BSAS Based Estimation

For time-varying nature of PHN, periodical pilot symbols are needed in each OFDM symbol at the cost of system bandwidth. This motivation leads this research to develop a blind algorithm to estimate PHN. The BSAS algorithm performs a single pass on a given data set. In addition, each cluster is represented by the mean of the vectors that have been assigned to it. For each new vector s, presented to the algorithm, its distance from the already formed clusters is computed. If these distances are larger than a (user-defined) threshold of dissimilarity,Θ, and if the maximum allowable number of clusters, q, have not been reached, a new cluster containing s is created. Otherwise, s is assigned to its closest cluster and the corresponding representative is updated. The algorithm terminates when all data vectors have been considered once. The BSAS algorithm can be expressed as follows:

Let
$m = 1$
$C_m = s_1$
for $i = 2$ to N,
Find $C_k : d(s_i, C_k) = min_{1 \le j \le m} d(s_i, C_j)$
If $(d(s_i, C_k) > \Theta) AND (m < q)$,
$m = m + 1$
$C_m = s_i$
else
$C_k = C_k \cup s_i$
end
end

To find out the angle of rotation caused by CPE as shown in Fig. (4), the received frequency-domain signal points were divided into groups according to the constellation size. From this group of signals, average angle w.r.t reference axis is estimated. The average angles estimated for all groups are expressed by

$$\phi_{av}(p) = \frac{1}{N_s}\sum_{j=1}^{N_s} arg\{C_j\} \qquad (12)$$

where, $p = 1, 2, ...M$ be the constellation size. The averaged angles were compared with the corresponding ideal angles that gives M individual CPE angles,

$$\phi_{cpe}(p) = \phi_{ideal}(p) \sim \phi_{av}(p) \qquad (13)$$

These CPEs were again averaged to figure out the desired angle of rotation and given by the following cost function,

$$\hat{\phi}_{proposed} = \frac{1}{M}\sum_{j=1}^{M} \phi_{cpe}(j) \qquad (14)$$

V. Simulation Results

To investigate the performance of the proposed method, several computer simulations were performed by using the following system parameters; The total number of sub-carriers, $N_s = 1024$, the data sub-carriers,$N_D = 824$ and unused sab-carriers were 200. The data symbols were modulated by using 64-QAM. The known channel type was 10 tap Rayleigh fading multipath channel. the sampling frequency and 3dB bandwidth were 20MHz and 10kHz respectively. The rms value of PHN was 3^0. The simulation results are depicted in Fig. (5) and (6). In Fig. (5) the transmit and received symbols after estimation are compared and it is clearly seen that most of the symbols are decoded efficiently, whereas in Fig. (6) bit error rate performance of proposed method is compared with conventional pilot based method.

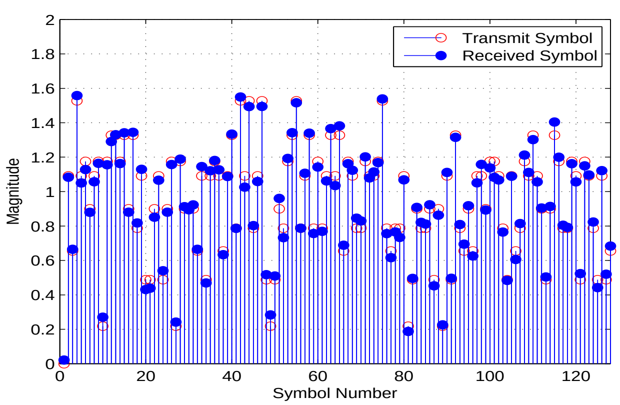


Fig. 5. Comparison of Transmit and Received Symbols.

VI. Conclusion

Though blind algorithm suffers from calculation complexity, to save the valuable bandwidth with the cost of lower calculation load an improved and bandwidth efficient new

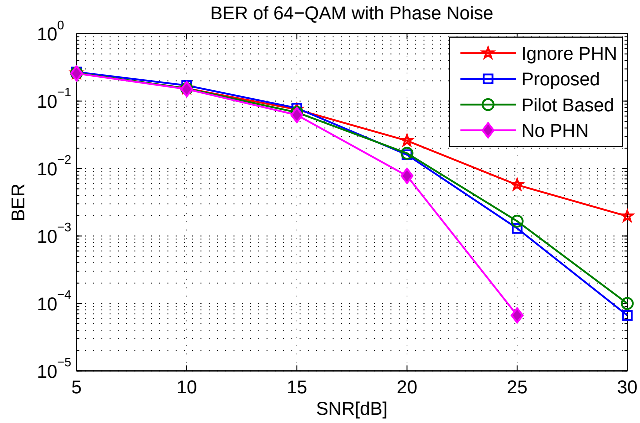


Fig. 6. BER Performance Comparison

algorithm has been offered for CPE compensation in OFDM systems. Simulations show that the proposed Basic Sequential Algorithmic Scheme smash the conventional pilot based CPE compensation method.

References

[1] Q. Zou, A. Tarighat, and ali H. Sayed, "Joint Compensation of I/Q Imbalance and Phase Noise in OFDM Wireless Systems," *IEEE Trans. on Communication.,* vol. 57, no. 2, Nov. 2009.

[2] P. Rabiei, and N. Won,"Reduced-Complexity joint baseband compensation of phase noise and I/Q imbalance for MIMO-OFDM systems," *IEEE Trans. Wireless Commun.,* vol. 9, no. 11, Nov. 2010.

[3] M. K. Lee, K. Yang, and K. Cheun,"Iterative receivers based on subblock processing for phase noise compensation in OFDM systems," *IEEE Trans. Commun.,* vol. 59, no. 3, Mar. 2011.

[4] G. Sridharan and T. J. Lim, "Blind Estimation of Common Phase Error in OFDM and OFDMA," *Proceedings of the IEEE GLOBECOM*, 2010.

[5] Y. Tsai, C. P. Yen, X. Wang, "Blind Frequency-Dependant I/Q Imbalance Compensation for Direct-Conversion Receivers," *IEEE Transactions on Wireless Communications*, vol. 9, no. 6, June 2010.

[6] M. K. Lee, S. C. Lim and K. Yang, "Blind Compensation for Phase Noise in OFDM Systems over Constant Modulus Modulation," *IEEE Transactions on Communications*, 2012.

[7] J. Tubbax, C. Boris, L. V. Perre, S. Donnay, M. Engels, H. De Man, and Marc Moonen, "Compensation of IQ Imbalance and Phase Noise in OFDM Systems," *IEEE Transactions on Wireless Communications*, vol. 4, no. 3, May 2005.

[8] Theodoridis, Pikrakis, Koutroumbas, Cavouras, "Introduction to Pattern Recognition: A Matlab Approach, 1st Edition," *Academic Press*, 2010

[9] G. Sridharan, "Phase Noise in Multi-Carrier Systems, " *Master Thesis*, University of Toronto, Canada, 2010.

[10] D. Petrovic, W. Rave, G. Fettweis, "Effects of Phase Noise on OFDM Systems With and Without PLL: Characterization and Compensation," *IEEE Transactions on Communications*, vol. 55, no. 8, August 2007

[11] F. Munier, T. Eriksson and A. Svensson, "An ICI Reduction Scheme for OFDM Systems with Phase Noise over Fading Channel," *IEEE Transactions on Communications*, vol. 56, no. 7, July 2008.

[12] M. Webster and M. Seals. Suggested phase noise model for 802.11hrb. [Online]. Available: https://mentor.ieee.org/802.11/dcn/00/11-00-0296-01-00sb-suggested-phase-noise-model-for-802-11-hrb.ppt

[13] M. A. Hossain; H. Lin; K. Yamashita,"Low-Complexity Blind Phase Noise Compensation in OFDM Systems," *Proceedings of ICEE-ICT*,Dhaka, 2014

Design and Implementation of a Painter Robotic Arm With Graphical User Interface

Mohammed Saifuddin Munna, Bijoy Kumar Tarafder, Md. Golam Robbani, Tuton Chandra Mallick
Department of Electrical and Electronic Engineering
Premier University, Chittagong
Chittagong- 4203, Bangladesh.
munna.puc@gmail.com, bijoytarafder@gmail.com , robbani.puc@gmail.com, tuton.soc@gmail.com

***Abstract*—Recent research on robots has been trying to develop intelligent robots that can match human behavior on high level intelligent tasks that require sensing, complex motion and intelligence. In the recent few years robot is given some artistic behavior that robot can sing, dance even robot can play games. This paper presents a control method for a robotic arm to let the robot to acquire another human artistic behavior "drawing". In the proposed design MATLAB has been used for image processing interface and trajectory calculation. Value of co-ordinate position of a pixel of binary image is converted into joint angle applying inverse kinematics and servos are controlled with Arduino Mega. Sobel and Canny are applied for edge detection method and authors showed comparison between the output images after applying Sobel and Canny edge detection method. Image is given input through Matlab GUI, which is user interactive.**

Keywords—Matlab; GUI; Robot Kinematics; Edge Detection; Image Processing, Arduino Mega

I. INTRODUCTION

Robotic industry has became a giant sector in the recent world. Scientists and technologists have been trying to add different dimensions in robot behavior. Modern developments on robotics have different types of applications in a wide range of industries, including education, health care, household services, military, entertainment, etc. Drawing is an art which requires high degree of artistic and innovative mentality of human mind. So it is a complex and challenging task to habituate a robot with this innovative art. Our research work is to add a different shade of robotic behavior in artistic manner that is eye catching, human friendly and inspiration to the general people to know about the robotic activities. Basically, Sketcher Robotic Arm is a 2 degree of freedom robotic arm that can sketch human face. It performs few basic actions of robotic control system that were intended to design with highly possible precise way with the locally available equipments. We built a Matlab Graphical User Interface to take the input image. Image is filtered with several filtering tools to make it a binary image which is suitable for the robot to understand.

II. RELATED WORK

There have been few researches on the human portrait generation system. Some researchers have conducted the study of drawing robots. ZKM laboratory created the first real-time robot portraitist system, in which an industrial robot drew the face portrait for the human sitting in front of the camera [7]. Switzerland LASA-EPFL laboratory developed the most complete cartography robot system integrating Fujitsu's humanoid robot HOAP-2 with a new image process technology to draw the portrait [1]. Reference [2] presents a behavior-based robot control method of brush drawing where differential movement was adopted instead of traversal points. Reference [3] presents a 3 DOF robotic arm used for drawing on a paper sheet Which is constructed using LEGO NXT bricks. The research in [1] aims to develop a human portrait generation system that enables the two-armed humanoid robot, Pica, to autonomously draw the face portrait of the person sitting in front of Pica. Betty, a portrait drawing robot was developed using modified Theta-graph, called Furthest Neighbour Theta-graph [4].

The different controlling methods were applied in the stated references. Some of them are too complex, some are expensive, some are extensive, although every research tried to find the finest way to draw. We wanted to make the robot free of complexity that's why we used 2 DOF arm. We tried to give the robot an easy interface and scope to be modified easily, that's the reason of using Matlab and it's GUI .

III. SYSTEM DESIGN

A. Working Procedure

The block diagram of working procedure is given in Fig.1. Image is taken input to the system through Matlab GUI. Image can be acquired by a camera connected to serial port or can be picked from a specified path in computer. The taken image is RGB, so the system converts the image into grayscale. Then the grayscale image is converted to a binary image by edge detection method. Binary image is nothing but a combination of black and white pixels. Matlab calculates the necessary inverse kinematics calculation taking black pixel's coordinate position and sends joint angle for the servos to the controller. Arduino receives the joint angle through serial communication and controls the servos.

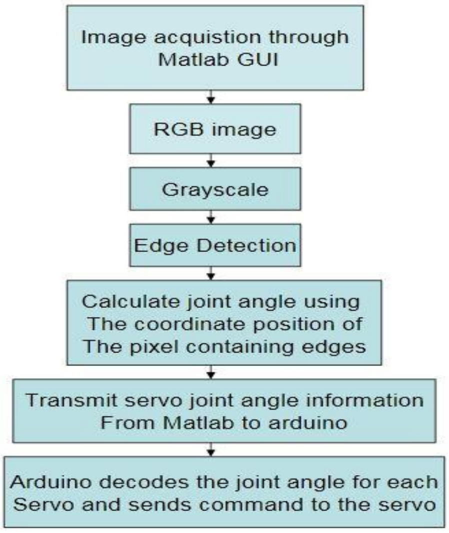


Fig.1. Block Diagram of Working Principle

B. *Robot Arm*

The designed arm is basically a 2 DOF planer robotic arm placed on a wooden board. The arm consists of two links made of aluminum sheet, one servo mount, two servo motor and a pen holder as an end effector. A suitable length of arm is maintained so that it can sketch on a A4 size paper sheet . Here end effector is a pen holder attached with a servo motor.

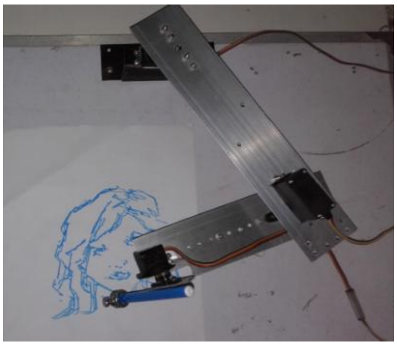


Fig.2(a). Robot drawing image

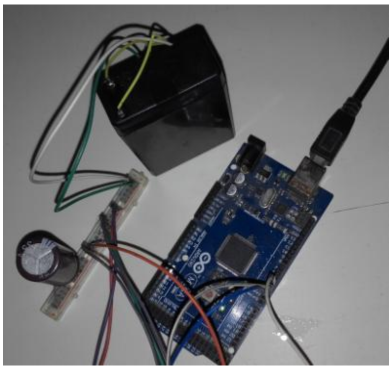


Fig.2(b). Controller Section

During line drawing pen is kept down and after drawing a line pen is kept up.

C. *Matlab Interfacing*

Matlab is an interactive software system for numerical computations and graphics. It is specially designed for matrix

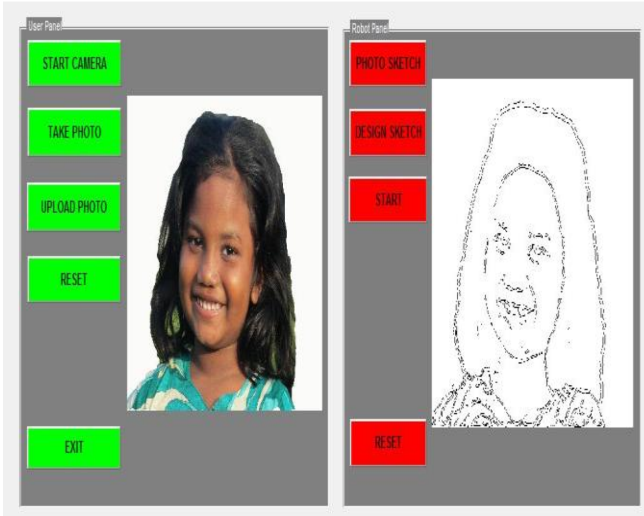


Fig.3. Matlab GUI

Computations, solving systems of linear equations. Matlab has a very large database of built in algorithm for image processing and computer vision applications. It provides many functions for image processing and other tasks. Most of these functions are written in Matlab language and are publicly readable that can be customized. Matlab can write serial data to serial port in different baud rate which are arduino readable. Serial communication between arduino and Matlab is a lot easier than any other system. Because of these advantages we choosed Matlab for image processing and inverse kinematics calculation.

Due to easy access to the image representing matrix Matlab makes it easy to apply the inverse kinematics to translate the pixel coordinate information to the robot joint angle. There is another fantastic part in Matlab, which is Graphical User Interface (GUI). Graphical User Interface can be designed easily in Matlab. GUI gives an user advantages to interact with the system architecture.

Fig.3 is our designed Matlab GUI which has been used to give an image as input. GUI has two input feature, acquiring an image by a web cam or from a folder inside computer.

D. *Arduino Mega*

Arduino is an open source physical computing platform based on a simple i/o board and a development environment. The Arduino Mega 2560 is a microcontroller board based on the atmega2560. Arduino Mega has been used for servo motor controlling. Arduino receives joint angle information as a serial data. After decoding the joint angle data arduino sends command to the servo how much should it rotate.

E. Servo Motor

Servo motor is an actuator that can be controlled precisely for linear or angular position. A servo motor consists of electric motor, feedback device and electronic controller. Servo is needed to feed a signal pulse to rotate for a particular angle. We used MG 996 servo motor which is high torque servo. To use the servo, firstly it needs to be calibrated. For 544us pulse MG 996 servo stays at its zero position and for 2400us that servo stays at 180 degree position.

IV. INVERSE KINEMATICS

Inverse kinematics is the method for determining the joint angle when the position of the end effector is known. Tasks to be performed by a manipulator are in Cartesian space whereas actuators work in joint space. The conversion of the position

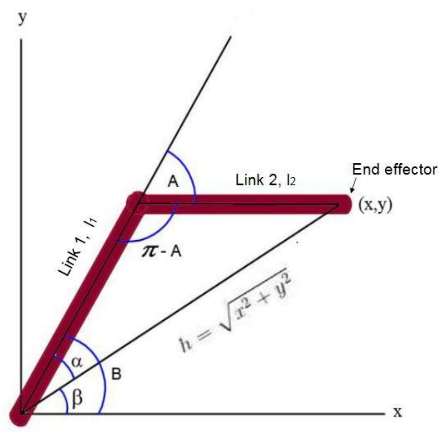


Fig.4. Geometry of Inverse Kinematics

and orientation of a manipulator end-effector from Cartesian space to joint space is called as inverse kinematics problem.

Let's establish few expressions to solve the inverse kinematics problem. In Fig.4, link 1 and link 2 forms a triangle with the joining line of the starting point of link 1 and ending point of link 2. So we can apply few equations from the concept of geometry. B is the angle of link 1 and A is the angle of link 2.

So, we can write,

$$\cos(\pi - A) = \frac{l_1^2 + l_2^2 - h^2}{2l_1 l_2} \quad (1)$$

(2)

$$\pi - A = \cos^{-1}\left(\frac{l_1^2 + l_2^2 - h^2}{2l_1 l_2}\right) \quad (3)$$

$$A = \pi - \cos^{-1}\left(\frac{l_1^2 + l_2^2 - h^2}{2l_1 l_2}\right)$$

A is the desired angle for link 2. The reference line for link 2 to form angle is the line parallel to the link 1. Now we will go for the angle of link 1.

$$\tan\beta = \frac{y}{x}$$

$$\beta = \tan^{-1}\frac{y}{x}$$

$$\cos\alpha = \frac{l_1^2 + h^2 - l_2^2}{2l_1 h}$$

$$\alpha = \cos^{-1}\frac{l_1^2 + h^2 - l_2^2}{2l_1 h}$$

Now, $B = \alpha + \beta$

So, if we know the length of the arm and the position of the end effector, we will find how much the servo motor should be rotated to reach the end effector in desired position.

V. IMAGE PROCESSING AND EDGE DETECTION

Image processing is one of the vital segments of the designed system. The input image is a RGB image which a combination of millions of shed that is too complex to understand for a robot. So we need to convert an RGB image to grayscale where different color combinations are reduced to only black and white intensity. Then grayscale image is converted into a binary image. The process of converting a grayscale image into a binary image is performed by edge detection method. Edges characterize boundaries and are therefore a problem of fundamental importance in image processing. Edge detection significantly reduces the amount of data and filters out useless information.

Few consecutive steps must be followed to implement the canny edge detection algorithm. The following steps are followed to perform canny edge detection. Firstly smooth the image with a Gaussian filters. Then compute the gradient magnitude and orientation using finite difference approximations for the partial derivatives. Then apply non maxima suppression to the gradient magnitude. Then use the double threshold algorithm to detect and line edges.

The sobel edge detection algorithm use two masks with 3*3 sizes, one estimating the gradient in the x-direction and the other estimating the gradient in the y-direction. The mask is slid over the image, manipulating a square of pixels at a time. The algorithm calculates the gradient of the image intensity at each point and fives the direction to increase the image

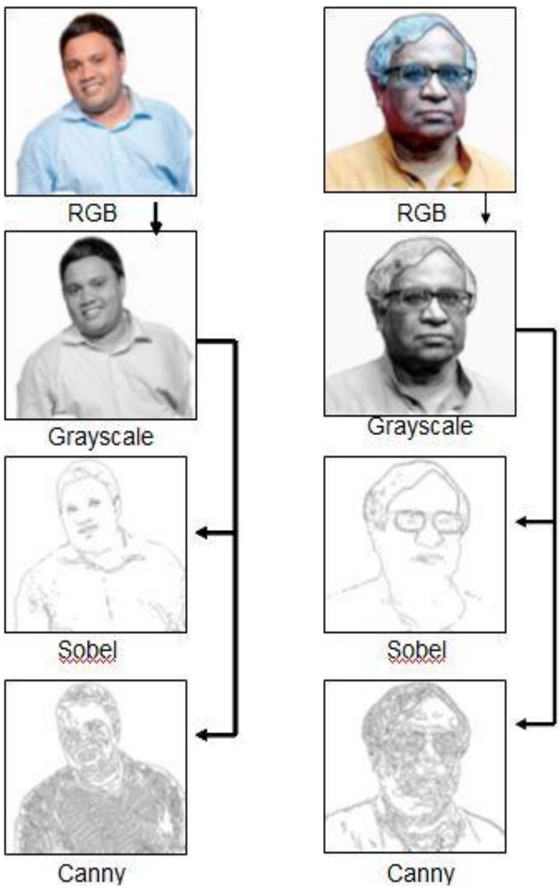


Fig.5. Canny and Sobel Edge Detection

intensity at each point from light to dark. Edges areas represent strong intensity contrasts which are darker or brighter.

Canny and sobel both are popular edge detection method but their working procedures are different. In Fig.5 we showed binary images after executing Canny and Sobel.

VI. PATH DRAWING ALGORITHM

A binary image is a combination of black and white pixel. When an edge detection method is applied an image is turned into a binary image where few lines or paths indicates the input image. Every line in the binary image contains black pixels attaching one with another. In Matlab a binary image represents a 2D matrix where every element contains the information of the pixel. White pixel represents 1 and black pixel represents 0. The first thing to draw a line is to detect the line from the edges after performing edge detection method. The basic idea is to search the image containing matrix for the black pixel from the first element of the matrix. Whenever a black pixel is found that is the beginning of the line and then path drawing algorithm is applied. Path drawing algorithm is a simple algorithm where the system search for the availability of a black pixel around another black pixel and calculate the joint angle from the black pixel coordinate value through inverse kinematics equations and transmit the joint angle

1	2	3	4	5	6	7	8	9	10
11	**12**	**13**	14	15	16	17	18	19	20
21	22	23	**24**	25	26	27	28	29	30
31	32	33	**34**	35	36	37	38	39	40
41	42	43	**44**	45	46	47	48	49	50
51	52	53	54	**55**	56	57	58	59	60
61	62	63	64	65	**66**	**67**	68	69	70
71	72	73	74	75	76	77	**78**	79	80
81	82	83	84	85	86	87	88	**89**	90
91	92	93	94	95	96	97	98	99	100

Fig.6. Neighbour Pixel Searching Method

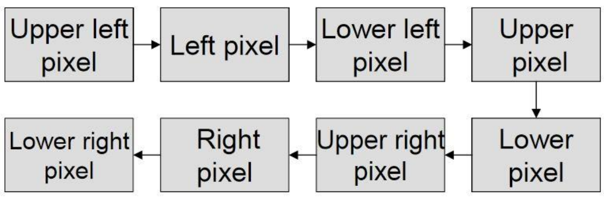


Fig.7. Pixel Searching Precedence

information to the controller to control the servo . Let it be described with the Fig.6.

Fig.6 represents a binary image after performing edge detection. There is a single edge consists of black pixels touching one after one makes a chain. The system starts to search from the first pixel if there is a black pixel available and the system finds it in 12 number element of the image containing matrix. Now the system knows that 12 number pixel is the beginning of the edge. Then the system calculate the joint angle from the coordinate value of the 12 number pixel and transmit the joint angle to the controller to reach the end effector in that coordinate point position. Then it searches around the 12 number pixel i.e. 1, 2, 3, 13, 23, 22, 21, 11. If there is a black pixel available and 13 number element is found as black pixel and transmits it joint angle information to the controller. Again the system searches around the 13th pixel for a black dot and so on till the last black pixel of that line.

There is followed a procedure (Fig.7) to search a pixel by its position priority. First, a pixel searches around itself for upper left pixel following by left, lower left, upper, lower upper right, right, lower right.

VII. Results and Discussions

That robotic arm sketched few image of different peoples. After the thorough design of the robot hardware, we obtained a structure that was mechanically strong and got a stable control. Edge detection was performed by Canny and Sobel, In Fig.8 there are few sketches by robot.

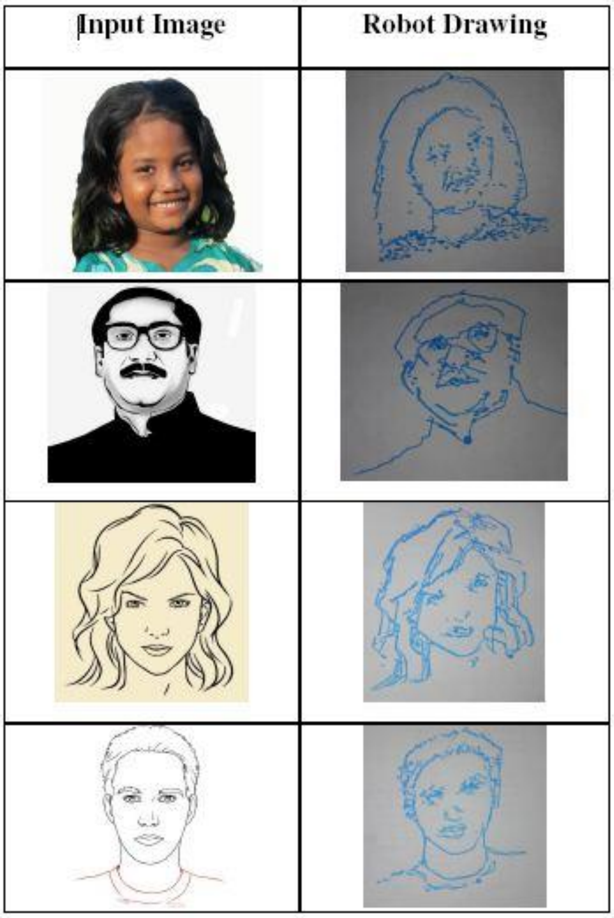


Fig.8. Input Image and Output Image drawn by robot

both of method has it's own advantages and limitations. Although Canny is an efficient method for edge detection in the sense of keeping the characteristics of input image but it picks a lot of edges which is tedious and time consuming to perform for a robot. On the other hand, Sobel detects less edges that makes robot faster to perform but sometimes it can not represent the input image due to the lack of enough edge.

VIII. Conclusion

In this paper we developed a robotic arm which can sketch human face with a 2 DOF robotic arm. Our research work will motivate the peoples about robots through an interesting behavior of robot. Our designated system has few advantages such as,

- It is a cost effective, frugality of complexity and user friendly robot.
- Graphical User Interface gives user more interaction to control the robot.
- Capable of producing higher quality output depending on efficiency of edge detection.

Our designed robot has a versatile application in entertainment and educational purpose. It's a great fun to watch that a robot is sketching image. It's an inspiration to the general people to know about the robotic activities.

Our future research is to develop the drawing quality by increasing the edge detection efficiency. Our target is to modify the current edge detection method to get finest drawing with less possible amount of edges.

IX. Reference

[1] Chyi-Yeu Lin, Li-Wen Chuang, Thi Thoa Mac, "Human Portrait Generation System for Robot Arm Drawing", 2009 IEEE/ASME International Conference on Advanced Intelligent Mechatronics, Singapore, July 14-17, 2009

[2] Yang Junyou, Qiu Guilin, Ma Le, Bai Dianchun, Huang Xu, "Behavior-based Control of Brush Drawing Robot", 2011 International Conference on Transportation, Mechanical, and Electrical Engineering (TMEE), Changchun, China, Dec. 2011.

[3] A´ kos Ha´mori, Ja´nos Lengyel, Barna Resko, "3DOF drawing robot using LEGO-NXT", 15th International Conference on Intelligent Engineering Systems, Slovakia, June 2011.

[4] Meng Cheng Lau, Jacky Baltes, John Anderson and Stephane Durocher, "A Portrait Drawing Robot Using a Geometric Graph Approach: Furthest Neighbour Theta-Graphs"

[5] Shunsuke Kudoh, Koichi Ogawara, Miti Ruchanurucks, Katsushi Ikeuchi, "Painting Robot with Multi-Fingered Hands and Stereo Vision", 2006 IEEE International Conference on

[6] Multisensor Fusion and Integration for Intelligent Systems, Germany, Sept. 2006.

[7] Sylvain Calinon, Julien Epiney and Aude Billard, "A Humanoid Robot Drawing Human Portraits", Ecole Polytechnique F´ed´erale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland

[8] Robot-lab Art, http://www.robotlab.de/index.htm

[9] Ahmed El-Barkouky, Ali Mahmoud*, James Graham and Aly Farag, "An interactive educational drawing system using a humanoid robot and light polarization", ECE Department University of Louisville, Louisville, KY 40292, USA

[10] Raman Maini & Dr. Himanshu Aggarwal, "Study and Comparison of Various Image Edge Detection Techniques", International Journal of Image Processing (IJIP), Volume (3) : Issue (1)

[11] Xiaoyang Mao, Yoshiyasu Nagasaka and Atsumi Imamiya, "Automatic Generation of Pencil Drawing From 2D Images Using Line Integral Convolution," *CAD/Graphics*, 2001.

[12] L. Chyi-Yeu, T. T. Mac, and C. Li-Wen, "Real-time artistic human face portrait by humanoid robot," in 2009 IEEE International Conference on Control Applications (CCA), Piscataway, NJ, USA, pp. 205-210, July 2009.

Gesture Based Wireless PC Control with Gyroscope and Accelerometer

Sabrina Tarannum[1], Mohammed Saifuddin Munna[2], Mohammad Ariful Islam Bhuyan[3]
Department of Computer Science & Engineering[1], Department of Electrical and Electronic Engineering[2, 3]
Premier University, Chittagong
Chittagong- 4203, Bangladesh.
saaabrin@gmail.com[1], munna.puc@gmail.com[2], arif.ajtfs@yahoo.com[3]

Abstract— **This paper proposed a design to control a PC with the virtual PC mouse by hand gesture. There are different types of methodology used to control virtual PC mouse i.e. kinetic sensor, webcam, optical sensor etc. But in our proposed design author used just an IMU, which is an inertial measurement unit. IMU combined with accelerometer and gyroscope. Firstly accelerometer and gyroscope track different gesture of the human hand. Then instruction sends over Bluetooth protocol. Inertial displacement data scaled for a frame and helps to move the pointer in different vector position. A JAVA programmed application is used in the Computer to perform the operation of Mouse. The total system and results are discussed later.**

Keywords-virtual; accelerometer; gyroscope; gesture; inertial measurement unit; JAVA programed ; bluetooth

I. Introduction

The use of computer makes human life easier. We cannot imagine our modern life without computer and its applications. In human computer interaction mouse is still commonly used input device. Now to control a PC mouse improved day by day. Many researchers used different methods to control PC mouse wirelessly. With rapid advancement in the field of human computer interaction, it has become possible to gain easy access and control of computer applications using gestures. The 2D mouse with three degrees of freedom can't satisfy their demands. They hope the mouse can control the rotations around the axes besides the movements along the axes.

Our proposed design based on accelerometer and gyroscope technology. We designed an artificial algorithm which helps to track the motion of hand and its gesture. Wireless device is interfaced over Bluetooth protocol.

On the other hand the receiver section has a JAVA programmed startup application. Which is used to auto start the program when the computer startup. Then JAVA program detect the data from sender & calculate it for mouse pointer position change and clicking option.

Moreover our proposed system is very cheap and user friendly.

II. Related Works

The hand and fingertips detection have been studied for several decades, for which methods designed for gesture recognition without any webcam and complexity.

In 2008 PC mouse operations using mouth open/close motions and head tilting is designed where a head sensor system is used . From the sensor, they collect the data from head movement angel & from this the roll & pitch is calculated. And a distance sensor is also placed in front of the mouth to calculate the mouse position.[3]
Another design for this types of device use gyro opto sensors for the operation where Optical IR sensor used to calculate the eye blinking & gyro calculate the body movement. [4]
Another work on this particular field works with image processing with the help of kinetic sensor. Here the movement of hand calculated for displacement & click also. [5]

III. System Overview

Our total system designed with two parts. One is remote device which is designed, and another one is PC processing unit. Remote part captures the hand gesture and motion. After that data processed by microcontroller then sends to PC unit over Bluetooth. PC unit have processing unit and then make decision to move the pointer in desire position according to hand gesture.

IV. Methodology

The whole system divided in two sections which are described below

A. Block Diagram

In this Proposed design there are two different sections first one is transmitter section and Second one is receiver section.

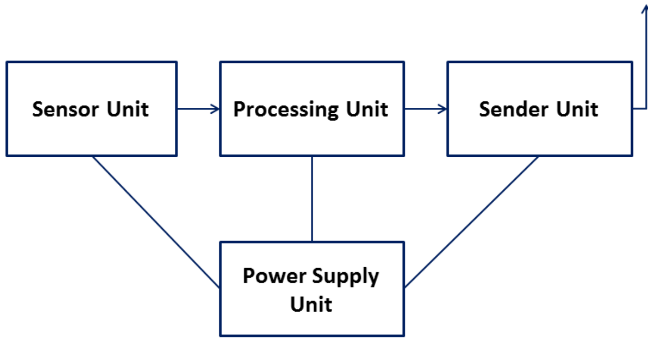


Figure 1. Block Diagram of Wireless Transmitter Section.

From figure 1. there four parts in transmeter section .

- Power suply unit
- Sensor Unit
- Processing Unit
- Sender Unit

In power suply unit it's a simple DC power suply of 5 volt.

In the sensor unit author used 2 differnet types of sensor for this proposed device.

- Accelerometer
- Gyroscope

Here 2 accelerometer is used as right click & left click option. And by using a single gyroscope this work is done.

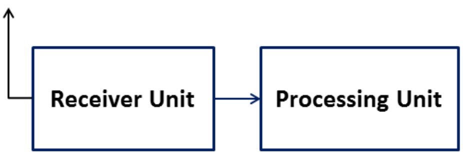


Figure 2. Block Diagram of Receiving Section.

On the other side the reciever unit have two differnt sections

- Receiver unit – Bulethoot
- Processing unit

This system use a JAVA based program as processiing unit of the resciever section.

B. *Hardware Unit*

- HC-05 Bluetooth Module
- Arduino UNO R3
- Gyroscope
- Accelerometer
- Flex sensor

a) *HC-05 Module:*

HC-05 module is an easy to use Bluetooth SPP (Serial Port Protocol) module, designed for transparent wireless serial connection setup. Serial port Bluetooth module is fully qualified Bluetooth V2.0+EDR (Enhanced Data Rate) 3Mbps Modulation with complete 2.4GHz radio transceiver and baseband. It uses CSR Bluecore 04-External single chip Bluetooth system with CMOS technology and with AFH(Adaptive Frequency Hopping Feature). It has the footprint as small as 12.7mmx27mm. Hope it will simplify your overall design/development cycle.[1]

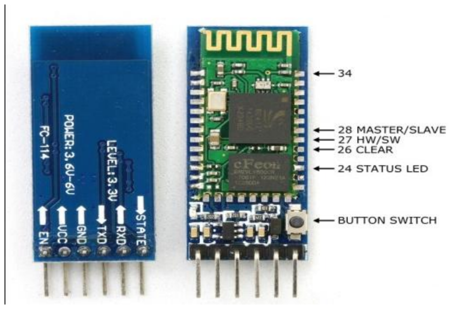


Figure 3. HC-05 module

b) *Arduino UNO R3 :*

It contains Atmega AVR 328p micro controller. It has 14 digital Output/Input pins (of which 6 can be used as PWM signals), 6 analog inputs, a 16 MHz crystal oscillator. Operating voltage of the micro controller is 5V. Contains 32 KB of flash memory enough to store the data for this particular project.[1]

In order to have proper mechanism and response from the robot , proper readings had to be taken for it. In this I have used 3 sensors:

Figure 4. Ardiuno Uno

c) *ADXL335 :*

It is a 3- axis accelerometer sensor, which can measures the forces applied on to the sensor in all the 3 directions X, Y and

Z axis. Further the raw data from the sensors are converted into acceleration by using some complicated equations. The advantage of the accelerometer was that the values do not change unless there is a change in position. But the problem with the accelerometer was that it contained high level of noise which makes the values inaccurate. So, to make these values accurate Gyroscope sensor was used.[1]

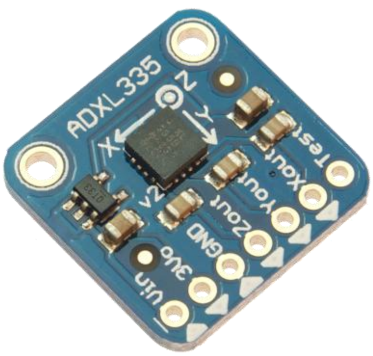


Figure 5. ADXL335 sensor

d) L3G4200D:

It is a 3- axis Gyroscope sensor which can measure the degree of rotation in all the 3 axis in form of alpha, beta and gamma. The values provided by the Gyroscope are very accurate but values do not remain static and tend to drift to the position Zero. To make the readings accurate as well as static both the values from the Accelerometer and Gyroscope were combined by using Kalman filters.[1]

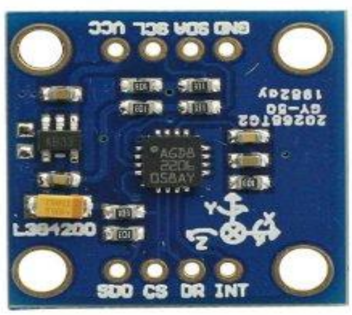


Figure 6. Gyroscope sensor

e) Flex Sensor :

Flex is basically a strip of carbon material having metal pads inside it. As the sensor is flexed, the resistance across the sensor increases. The resistance of the flex sensor changes when the metal pads are on the outside of the bend Straight (un-flexed) resistance: 9000 ohm, 90 degree bend resistance 14000 ohm and at 180 degree bend resistance 22000 ohms.[1] This resistance was calibrated and converted into angles and send the data of angel to the controller.

C. Calculaton of angel by gyrosocope & acclerometer

For calculating the amount of hand movement for cursor movement, some calculation has to be done from the raw data collected from the sensors.

Let, acceleration component denote as x_a ,y_a and z_a and gyro component denote as x_g , y_g and z_g.

If R vector on the XYZ axes. Please notice the following relation:

$$R^2 \mathrm{H}\, R_x{}^2 + R_y{}^2 + R_z{}^2$$

Since the tilt of the X-axis actually shows rotation around the Y-axis, and the angling of the Y-axis shows (negative) rotation around the X-axis.

The rotation around the X-axis (φ)

$$\varphi \approx \tan^{-1}\left(\frac{y_a}{\sqrt{x_a^2 + z_a^2}}\right)$$

The rotation around the Y-axis (ρ)

$$\rho \approx \tan^{-1}\left(\frac{x_a}{\sqrt{y_a^2 + z_a^2}}\right)$$

If the Z-axis is aligned along the gravitational acceleration vector, then it is impossible to compute rotation around the Z axis from the accelerometer. That's why need gyro values to find out Z rotational angle.

$$GyroXgyroX/131.0\ rate =$$
$$GyroY(gyroY/131.0)\ rate = \square$$

Where 131 is constant.

The rotational angle along with Z-axis can be define as

$$gyro\_\ angle\ \_\ z = (gyroZ_{rate} \times dt) + lastZ_{angle}$$
$$kalAngleX = (accX_{angle},\ gyroX_{rate})$$
$$kalAngleY = (accY_{angle},\ gyroY_{rate})$$

All of these equations are generalized form and shows the calculation methodology. To perform actual Kalman calculation library function needed. [2]

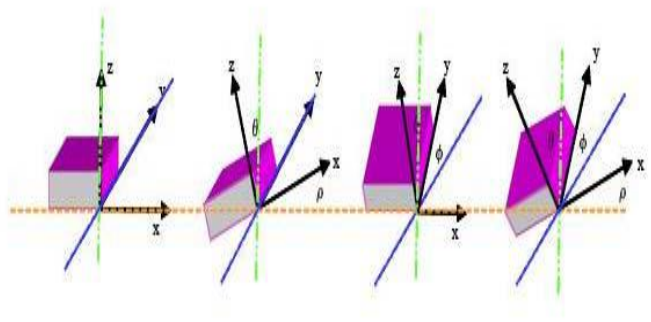


Figure 3. Geometry of angle calculation using accelerometer and gyrometer.[2]

D. Schemetic diagram of transmiter section:

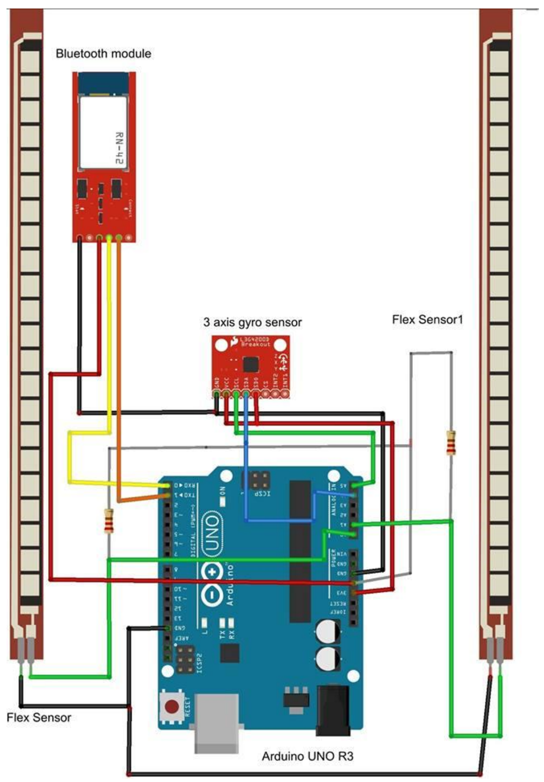


Figure 7. Schemetic Diagram for Transmiter Section

E. Algorithm

a) Transmitter section:

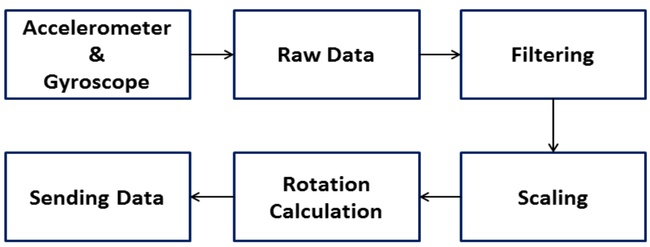


Figure 8. Processing Unit of Transmitting Section.

From the above figure 7, when the power of our proposed sytem is turn on then it starts to collect data from the sensors. Then the raw data is come to the ardiuno uno for proper calcuation and filtering. It calculate the hole plane a 3D plane of 800-1200-600 . Then it calculate the current positing of mouse and then calculate the position change and filter the data for our desire level. After calculating the position change dta it sends it to the reciver via bluetooth .

On the Other hand two different accelerometer sensors are palced into tow finger position . One for right click and the second one is for left click. When used click his 2nd finger it will send a 1 bit data to the controller & for middle finger it will produce a different bit. And when tow fibgers are moved at a same time it will work as double click.

In proposed system , the transmitting Controller creats differnts signal for every command. And using the HC-05 Module bluetooth Module it will send s to our reciever section.

b) Reciever Secsion:

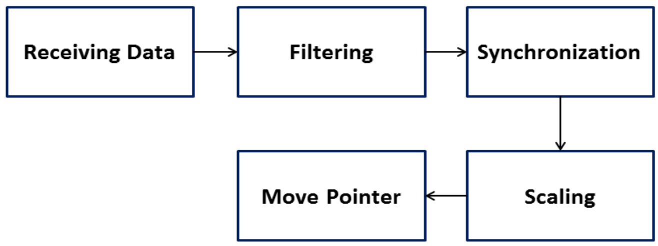


Figure 9. Processing Unit of Receiving Section.

In the this part from the flow chart shows in figure 8, we first recive a data via computer Bluetooth port or external bluethhot reciver.

At first our designed JAVA language program set to startup manu for auto start the program when the computer starts. This JAVA program used ROBOT class for mouse pointer movement.

When it takes a data bit via bluetooth then processing section stars to work. Calculate the change of position , instruct the pointer to move to the desire position.

In case of Click ‘r’ cherecter data from sender is processed to work as right click & ‘l’ cherecter from sender work as left click. When the processor find two ‘l’ data with in 0.2s it will work as double click.

V. RESULTS

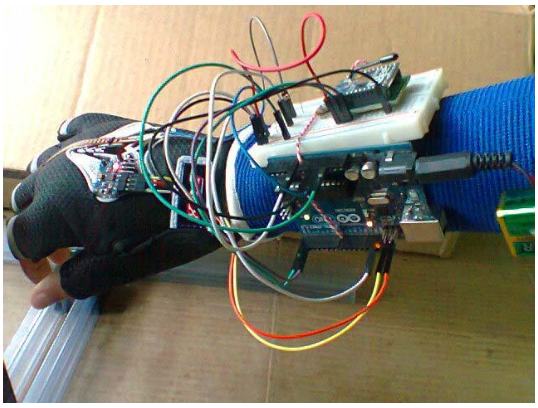


Figure 10. Designed transmission section.

Figure 9. show our designed hardware unit. Where 2 accelerometer & a gyroscope is connected with the transmiter bluetooth.

O n the other hand we collect some data from practical operation using our proposed design. We did it for different distances. bellow data table show the value of results.

TABLE 1: RESULT DATA

Observation no.	Distance from receiver (Meter)	Time required		
		Single click	Double Click	Cursor movement
1	8	1s	1s	(3-4)s
2	7	.5s	1s	(2-3)s
3	6	0.5s	1s	2s
4	5	No delay	0.5s	1s
5	3	No delay	No delay	(0.5 – 1)s
6	1	No delay	No delay	No delay

REFERENCES

[1] Shamsheer Verma “Hand Gestures Remote Controlled Robotic Arm,” Advance in Electronic and Electric Engineering. ISSN 2231-1297, Volume 3, Number 5 (2013), pp. 601-606

[2] Ariful Islam Bhuyan1, Tuton Chandra Mallick, “Gyro-Accelerometer based control of a robotic Arm using AVR Microcontroller”, 9th International Forum on Strategic Technology (IFOST), 2014.

[3] Nobuaki NAKAZAWA, Naoto HASHIMOTO and Kou YAMADA, “Welfare-Support Interface for PC Mouse Operations Using Mouth Open/Close Motions and Head Tilting”, Industrial Electronics, 2008. IECON 2008. 34th Annual Conference of IEEE

[4] Soochan Kim,Minje Park,Sasiporn Anumas, Jaeha Yoo, “Head Mouse System based on Gyro- and Opto- Sensors”, 3rd International Conference on Biomedical Engineering and Informatics (BMEI 2010)

[5] Szilvia Szeghalmy, Marianna Zichar,Attila Fazekas, “Gesture-based computer mouse using Kinect sensor”, 5th IEEE International Conference on Cognitive Infocommunications- 2014.

Analysis of Chirp Induced Impairments in Fiber Optic Transmission Systems Using Various Types of Fiber

Kanis Fatima[1], Md.Saiful Islam[2]
[1,2]Institute of Information and Communication Technology
Bangladesh University of Engineering and Technology
Dhaka,Bangladesh
Email:kanis.eee@diu.edu.bd

***Abstract*—Forthe high bit rate, large capacity multi-channel fiber communications the group velocity dispersion(GVD), non-linearity and loss effect has a great impact on the performance of fiber optic communication. On the other hand chirping plays an important role on pulse broadening. This paper represents an analytical model of pulse broadening factor with chirping and first order group velocity dispersion. For different values of chirping parameter the pulse broadening factor is analyzed for different bit rates and fiber parameters. Results showed that chirping has a significant impact on pulse broadening. For negative chirping there is no pulse compression and when the bit rate of the system increases the pulse broadens more with the fiber length and limits the system performance.Results showed that large effective area fiber (LEAF) shows better performance than standard single mode fiber (SSMF) and non-zero dispersion shifted fiber (NZDSF).**

***Keywords*—GVD,SSMF,LEAF,NZDSF**

I. INTRODUCTION

Group velocity dispersion (GVD) is one of the main factors limiting transmission length in high bit rate optical transmission system. When the fiber length is such that $L<<L_{NL}$ but $L\sim L_D$ then the pulse evolution is governed by group velocity dispersion .In this case the nonlinear effect is less [1].Most analytical expression showing the pulse broadening effect considering the Gaussian pulse envelope as input when considering first order GVD [2].The first order GVD effects with various chirp parameter is shown mostly [2].Analysis shown that a chirped Gaussian pulse broadens monotonically but at a rate faster than that of un-chirped pulse. In different dispersion regimes the un-chirped and chirped Gaussian pulse propagation comparison is shown[3].Most of the methods do not show the chirped Gaussian pulse behavior for various data rates with the different fiber parameters [4].One approach to evaluate the performance of optical fiber link is to analyze the pulse broadening caused by different factors and how to compensate this broadening effect [5].In this research paper, investigation has been carried out to find pulse broadening factor by solving nonlinear Schrödinger equation(NLSE),considering the first order GVD with the chirped Gaussian pulse propagation at various data rates as a function of transmission distance using different fibers. Next, for different fiber parameters the pulse broadening effectsare analyzed for various positive and negative chirping values at different data rates.

II. THEORITICAL BACKGROUND

If Gaussian input pulse is chirped prior to transmit it into the fiber then for the different chirp parameter they show interesting behavior when considering first and second order GVD. A chirped Gaussian pulse is mathematically represented by:

$$A(0,T) = \exp - (1+iC)\frac{T^2}{T_0{}^2} \qquad (1)$$

Where, C illustrates thechirp parameterwhichmay benegative orotherwise.Ifweset $C = 0$,this equation will reduce to un-chirpedGaussianpulse.

The fiberloss, GVDandnonlinearity effects occursimultaneously duringthepropagationofanopticalpulse.Theseeffectscan bemodeledinsinglemodefiber usingthe following nonlinear Schrödinger equation [1].

$$\frac{\partial A}{\partial z}+\frac{i}{2}\beta_2(z)\frac{\partial^2 A}{\partial T^2}-\frac{1}{6}\beta_3\frac{\partial 3_A}{\partial T^3}+\frac{\propto(z)}{2}A = i\gamma|A|^2A \quad (2)$$

WhereAis the slowly varying amplitude ofthe pulseenvelop; i is the imaginary vector,α is thefiberloss,$\gamma=\frac{n_2\omega_0}{CA_{eff}}$ represents the nonlinear effect where,$\omega_0$represents the carrier frequency, c is the velocity of light,$n_2$is the nonlinear refractive index and A_{eff} is the effective cross sectional area of the fiberandβ_2, β_3 represents the first and second order GVDeffects respectively;T is measured in frame of reference moving width the pulse at the group velocity$v_g(T = t - z/v_g)$ and z is the propagated distance. In equation (1),the effect of GVD isincludedthroughdispersion parameterthat'srelatedto$\beta_2$as$\beta_2=(-D\lambda^2)/(2\pi c)$;whereλis theoperatingwavelength, D is the dispersion parameterandcisvelocityoflightinfree space. If we use the normalized amplitude A(z,T) then we can write the linear partial differential equation in the form,

$$i\frac{\partial A}{\partial z}=\frac{\beta_2}{2}\frac{\partial^2 A}{\partial z^2} \qquad (3)$$

If $\tilde{A}(z,\omega)$ is the Fourier transform of A(z,T) then,

$$A(z,T)=\frac{1}{2\pi}\int_{-\infty}^{+\infty}\tilde{A}(z,\omega)\exp(-i\omega T)\,d\omega \quad (4)$$

The above equation satisfies an ordinary differential equation,

$$i\frac{\partial \tilde{A}}{\partial z} = -\frac{1}{2}\beta_2\omega^2\tilde{A} \quad (5)$$

whose solution is :

$$\tilde{A}(z,\omega)=\tilde{A}(0,\omega)\exp(\frac{1}{2}\beta_2\omega^2 z) \qquad (6)$$

This equation (6) shows that phase of each spectral component of the pulse changes due to GVD.Using equation (6) in (4) the general solution of (2) is,

$$A(z,T)=\frac{1}{2\pi}\int_{-\infty}^{+\infty}\tilde{A}(0,\omega)\exp(\frac{i}{2}\beta_2\omega^2 z - i\omega T)d\omega \quad (7)$$

Where ,$\tilde{A}(0,\omega)$ is the fourier transform of the incident field at z=0 and is found by,

$$\tilde{A}(0,\omega)=\int_{-\infty}^{+\infty}A(0,T)\exp(i\omega T)dT \quad (8)$$

By substituting equation (1) in (8) $\tilde{A}(0,\omega)$ is given by,

$$\tilde{A}(0,\omega)=\sqrt{\frac{2\pi}{1+1C}}\,T_0\exp\left[\frac{-\omega^2 T_0^2}{2(1+iC)}\right] \quad (9)$$

$$Here, a=\frac{1+iC}{2T_0^2}, b=-i\omega$$

As for a general quadratic exponent we know,

$$\int_{-\infty}^{+\infty}\exp[-(ax^2+bx+c)]dx=\sqrt{\frac{\pi}{a}}\exp(\frac{b^2-4ac}{4a}) \quad (10)$$

To obtain the launched field ,put $\tilde{A}(0,\omega)$ from (9) to (7).The integration can be performed analytically using (10),

$$A(z,T)=\frac{T_0}{\sqrt{T_0^2-i\beta_2 z(1+iC)}}\exp\left\{-\frac{(1+iC)T^2}{2[T_0^2-i\beta_2 z(1+iC)]}\right\}$$

Here, $a=\frac{T_0 2}{2(1+iC)}-\frac{i}{2}\beta_2 z$,b=iT,c=0

Considering the dispersion length $L=\frac{\beta_2 z}{T_0^2}$

The broadening equation considering the first order GVD with the chirped Gaussian pulse propagation is given by:

$$\frac{T_1}{T_0}=\sqrt{\left[\left(1+\frac{C\beta_2 z}{2\sigma_0^2}\right)^2+\left(\frac{\beta_2 z}{2\sigma_0^2}\right)^2\right]}$$

$$=\sqrt{(1+LC)^2+L^2}$$

In our research work we use the following parameters to find our desired result.

TABLE I. SIMULATION PARAMETERS FOR DIFFERENT FIBERS

Parameter(Unit)	**Fiber types**		
	SSMF	**LEAF**	**NZDSF**
Nonlinear Refractive Index,n_2(m^2/W)	2.35×10^{-20}	2.35×10^{-20}	2.35×10^{-20}
Wavelength,λ(nm)	1550	1550	1550
Input Power ,P(mW)	40-80	40-80	40-80
Bit Rate ,B(Gb/s)	20-50	20-50	20-50
Dispersion parameter, D(ps/nm-km)	17	3.5	5
First order GVD, β_2(ps^2/km)	-21.7	-4.46	-6.38
Effective area, A_{eff} (μm^2)	80	72	50
Chirp Parameter	Varied	Varied	Varied

III. RESULTS AND DISCUSSIONS

Fig.1 shows the analytical pulse broadening factor caused by first order GVD considering chirped Gaussian pulse in a standard single mode fiber (SSMF) optical transmission system ,operating in 1550nm wavelength at the data rates of 20Gbps and 50Gbps considering input power is 40mW.At 20Gbps data rate the pulse broadening factor is almost constant up to fiber length 1.4km for positive chirping, while at 50Gbps data rate the length is 1.24km.But for un-chirped pulse at 20Gbps the pulse broadening factor remains constant up to length 4.65km and 1.07km for 50Gbps bit rate. At bit rate 20Gbps and 50Gbps the corresponding half width of the pulse is 21.3ps and 8.5ps respectively. At this pulse widths the dispersion length, at which the pulse width becomes $\sqrt{2}$ times of the initial width defined by $L_D=\frac{T_0^2}{\beta_2}$is 20.9km and 3.33km respectively. At 20Gbps data rate, the pulse broadening factor becomes double at 36.54km while it is 4.86km at 50Gbps data rate for positive chirping. Again for un-chirped pulse it becomes double at 36.13km and 6.12km for 20 and 50 Gbpsdata rate respectively.

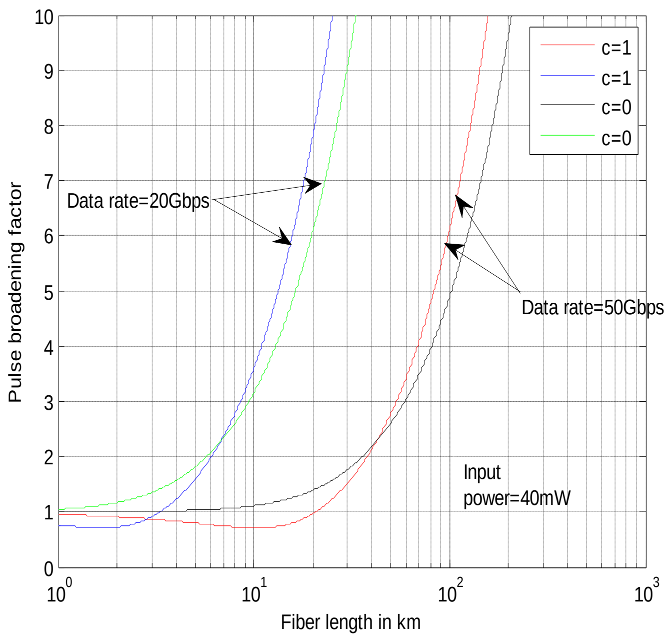


Figure1. Positive chirping effect inSSMF

Fig.2 shows for negative chirping at 20Gbps bit rate the initial pulse broadening factor is 1.05 and at 50 Gbps it is 1.334.

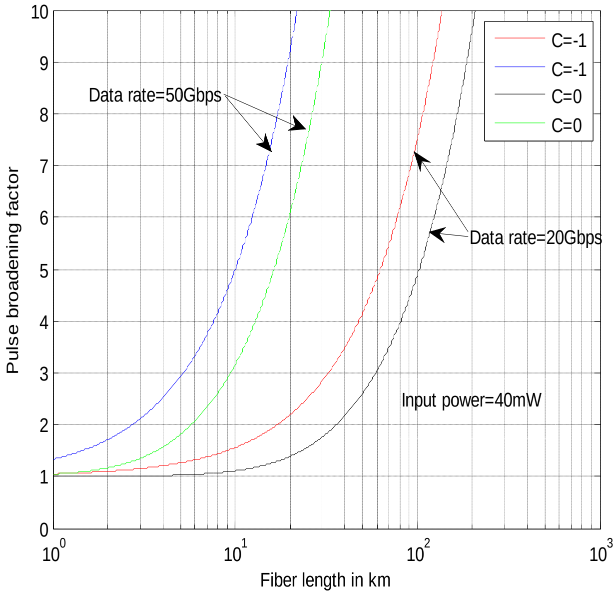


Figure 2. Negative chirping effect in SSMF

It is also seen that for negative chirping there is no pulse compression and with the fiber length the pulse broadening factor increases rapidly.The pulse broadening factor becomes double at 18.74km when the system operates at 20Gbps but it becomes 4.4km when the system operates at 50Gbps for negative chirping. Whereas for un-chirped pulse that is for c=0 at 20 Gbps bit rate the pulse broadening factor is almost constant up to distance 4.72km while at 50Gbps it is 1.5km.After this length the pulse broadening ration increases rapidly.

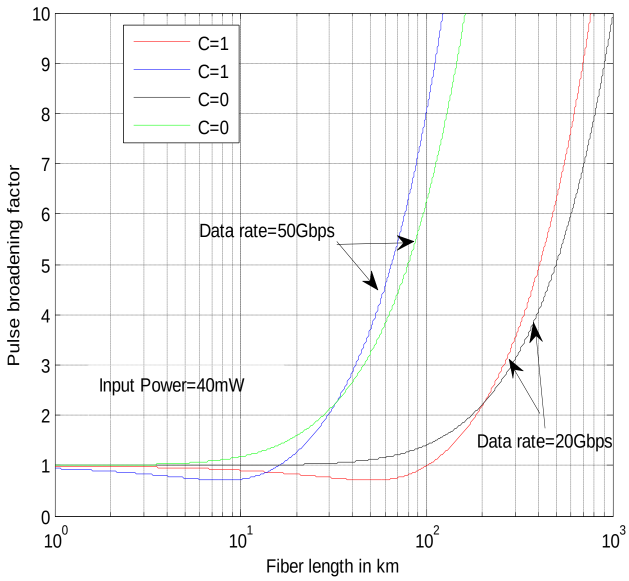


Figure 3. Positive chirping effect in LEAF

Fig. 3 shows the pulse broadening factor with the distance for LEAF fiber. From figure it is observed that at bit rate 20Gbps the pulse broadening rate is almost constant up to distance 3.72 km for positive chirping, while at 50Gbps it is 1.14km. At bit rate 20Gbps and 50Gbps the corresponding half width of the pulse is 21.3ps and 8.5ps. The corresponding dispersion length is 101.72km and 16.2 km respectively. The pulse broadening factor becomes double at 183.5km when the system operates at 20Gbps but it becomes 28.13km when the system operates at 50Gbps for negative chirping. For un-chirped pulse, the pulse broadening factor is almost constant up to distance 21.63km and 3.72km for bit rate 20Gbps and 50Gbps respectively. At 20Gbps bit rate the broadening factor becomes double at 175.8km and 28.12km at 50Gbps.

In Fig.4 for negative chirping at 20Gbps bit rate the initial pulse broadening factor is 1.01 and at 50 Gbps it is 1.063.It is also seen that for negative chirping there is no pulse compression and with the fiber length the pulse broadening factor increases rapidly. It is observed that at bit rate 20Gbps the pulse broadening rate is almost constant up to distance 2.34 km for negative chirping, while at 50Gbps it is 1.2km.The pulse broadening factor becomes double at 85.02km when the system operates at 20Gbps but it becomes 14.91km when the system operates at 50Gbps for negative chirping. Whereas for un-chirped pulse that is for c=0 at 20 Gbps bit rate the pulse broadening factor is almost constant up to distance 22km while at 50Gbps it is 3.6 km.After this distance the pulse broadening ratio increases rapidly.

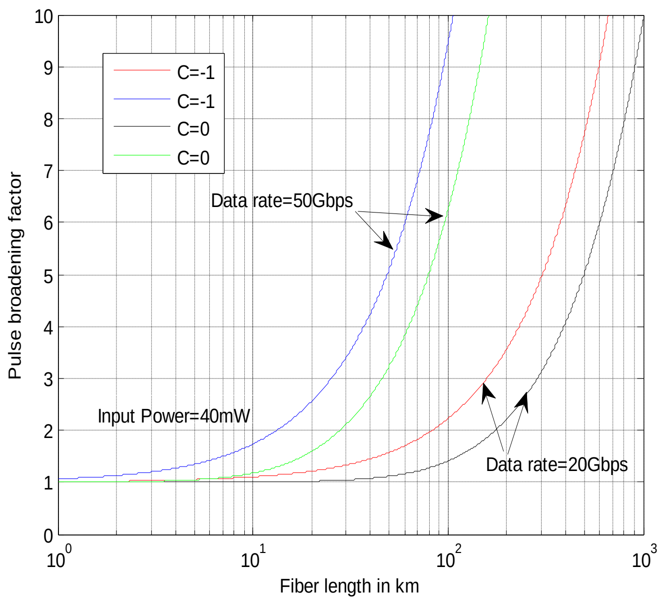


Figure4. Negative chirping effect in LEAF

Fig.5 shows the pulse broadening factor with the distance for NZDSF fiber. From figure it is observed that at bit rate 20Gbps the pulse broadening rate is almost constant up to distance 2.56 km for positive chirping, while at 50Gbps it is 1.07km.At bit rate 20Gbps and 50Gbps the corresponding half width of the pulse is 21.3ps and 8.5ps respectively. The corresponding dispersion length is 71.11km and 11.32 km respectively. The pulse broadening factor becomes double at 127.9km when the system operates at 20Gbps but it becomes 19.24km when the system operates at 50Gbps for negative chirping. For un-chirped pulse, the pulse broadening factor is almost constant up to distance 15.33 km and 3.5km for bit rate 20Gbps and 50Gbps respectively.

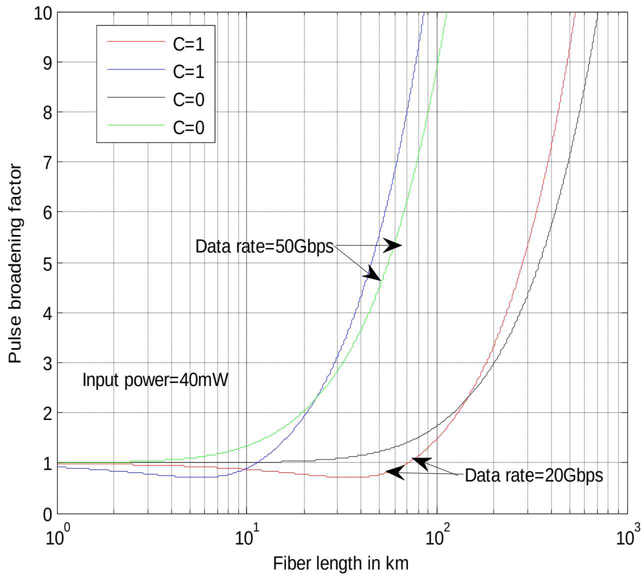


Figure5. Positive chirping effect inNZDSF

In Fig.6 for negative chirping at 20Gbps bit rate the initial pulse broadening factor is 1.014 and at 50 Gbps it is 1.112 in NZDSF.It is also observed that for negative chirping there is no pulse compression and with the fiber length the pulse broadening factor increases rapidly.

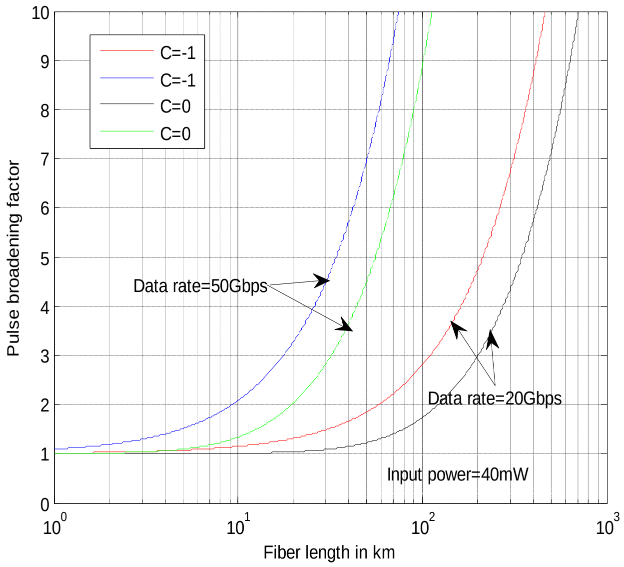


Figure 6. Negative chirping effect in NZDSF

It is observed that at bit rate 20Gbps the pulse broadening rate is almost constant up to distance 1.71 km for negative chirping, while at 50Gbps it is 1.21 km.The pulse broadening factor becomes double at 59.87km when the system operates at 20Gbps but it becomes 11.25 km when the system operates at 50Gbps for negative chirping. Whereas for un-chirped pulse at 20 Gbps bit rate the pulse broadening factor is almost constant up to distance 15.58 km while at 50Gbps it is 3.79 km. After this distance the pulse broadening ratio increases rapidly. At 20Gbps the broadening factor becomes double at 125.2km and 21.7km at 50Gbps.

IV. CONCLUSION

The combined effects of chirping and GVD on the transmitting pulse in optical transmission system have been investigated.For this an analytical model is developedconsidering chirped Gaussian pulse. Then the output pulse behavior at different bit rates, power and GVD parameters arevisualized graphically by MATLAB program.After analysis, the results showed that the impairments of chirping and GVD are more on SSMF than LEAF and NZDSF.It is also found that the negative chirping effect on pulses is more than positive chirping in all the 3 different types of fiber. It is also observed that input power has no effect on pulse broadening.

ACKNOWLEDGEMENT

This work has been carried out as a part of M.Sc. thesis work in the Institute of Information and Communication Technology of Bangladesh University of Engineering and Technology (BUET), Dhaka

REFERENCES

[1] G.P.Agrawal,NonlinearFiberOptics,4th ed., AcademicPress, 2008.

[2] M.Miyagi and S.Nishida,“Pulse Spreading in a Single-Mode Fiber Due to Third-Order Dispersion”, Applied Optics, Vol.18, No.5, March 1979.

[3] M. Y.Hamza, S. Farah, S. Yang and N. Sarwar, “Investigations of Un-chirped and Chirped Gaussian Pulse Propagation Through Single Mode Fiber in NormalVersusAnomalous Dispersion Regimes”,33rd International Conference on Telecommunication and Signal Processing-ISP2010.

[4]N. Sultanaand M.S.Islam, “The Effects ofCross-phase Modulation andThird Order Dispersionon Pulses and Pulse Broadening factorintheWDMTransmission System”, Proceedings of 2013 2nd International Conference on Advances in Electrical Engineering (ICAEE 2013),Bangladesh ,December, 2013.

[5]M.Y.Hamza,S.Tariq,M.M.AwaisandS.Yang, “Mitigationoftheeffectsofself-phase modulationandgroup-velocity dispersioninfiberopticcommunications:dispersion-andpower-map co-

optimizationusingthegeneticalgorithm",OpticalEngg.vol.47, No.7,USA,July2008.

[6]C.D.R.Campos, P.R. Horche and A.M. Minguez,"Interaction of Semiconductor Lase Chirp with Fiber Dispersion:Impact on WDM Directly Modulated System Performance",The Fourth International Conference on Advances in Circuits,Electronics and Micro-electronics,pp-17-22,2011.

[7]M. Amemiya, "Pulse Broadeng due to Higher Order Dispersion and its Transmission Limit",Journal of Lightwave Technology,Vol.20,No.4,April 2002.

The Promise and Challenges of Enhancing Solar Cell Efficiency Using Patterned Nanostructures

Saniat Ahmed Choudhury
Department of Electrical and Electronic Engineering
Independent University, Bangladesh
Dhaka, Bangladesh
Email: ac.saniat@gmail.com

Mustafa Habib Chowdhury
Department of Electrical and Electronic Engineering
Independent University, Bangladesh
Dhaka, Bangladesh
Email: mchowdhury@iub.edu.bd*

***Abstract*— This study provides a brief review of the different methods of using patterned nanostructures to improve light-coupling efficiency in Photo-voltaic (PV) cells. The different kinds of patterned nanostructures reviewed include nanowire solar cells with nanopillars placed periodically on top of a Si substrate, quantum dot solar cells, mesoscopic solar cells, and plasmonic solar cells. For plasmonic solar cells, the relationship between energy conversion efficiency in thin-film Si solar cells and the type of metallic nanoparticle, the size of the metallic nanoparticles, and the distance between neighboring metallic nanoparticles in a periodic array of the metallic nanoparticles deposited on top of the Si substrate are explored. Of the different metallic nanoparticle systems studied, it is found that silver nanoparticles with diameter of 100 nm and spaced 220 nm from neighboring nanoparticles in a periodic array of the metallic nanoparticles deposited on top of the Si substrate provides a significant increase in the generated short circuit current density (J_{SC}) in the wavelength region of λ = 400 – 1100 nm. Aluminum nanoparticles with diameter of 100 nm and spaced 220 nm from neighboring nanoparticles provides a significant increase in the generated short circuit current density (J_{SC}) in the relatively bluer wavelength region of λ = 200 – 600 nm. This can be attributed to the plasmon resonance wavelength of Ag being more red-shifted than that of Al. The increase in J_{SC} appears to be strongly correlated to the plasmon resonance wavelength of the metal nanoparticles. Additionally, this study also shows the near-field enhancement image for an Ag nanoparticle on top of a Si substrate. The near-field images suggest that enhancements in the near-fields can eventually lead to the increase in J_{SC} as shown in this study.**

Keywords- photovoltaics, plasmonics, thin-film solar cells, light trapping, nanoparticles, surface plasmons, silver nanoparticles, gold nanoparticles, aluminium nanoparticles.

I. INTRODUCTION

As we delve further into a future where our dependence on fossil fuels must be curtailed due to the non-renewable nature of fossil fuels, the interest in alternative and renewable sources of energy has been greatly peaked. Solar energy is the most promising of these alternatives. Photovoltaic (PV) cells are utilized to convert the energy from the Sun into electrical energy [1] – [3]. Theoretically, this should provide a potentially unlimited source of clean and sustainable electrical energy. However, two major complications prevent them from revolutionizing the generation of energy. Firstly, the conversion efficiency of commercially available PV cells leaves a lot a to be desired – at their most optimum conditions, the efficiency is no more than 20%. Secondly, the cost of PV cells is an issue. Although costs have decreased greatly over the past few decades, they are yet to be at a financially viable level [4]. The record efficiencies for Si based PV cells is 25%, while overall it is 41% for GaAs PV cells – both have too large a cost-to-efficiency ratio [4], and thus are commercially unfeasible. The challenge, therefore, is to reduce this ratio to as low a value as possible, as the possible benefits are immense.

The energy conversion efficiency depends on certain factors [5], including

- Optical absorptivity
- the minority carrier lifetime, which is the average time a carrier can exist in its excited state.
- diffusion length, which is the average length a carrier can travel before recombination.

To keep these factors in check, significant amount of research, including ours, has been focusing on thin-film PV cells (less than 10μm thick), as they reduce the mismatch between the electronic and photonic length scales [6]. However, the reduced thickness results in lower number of electron-hole pairs produced, and hence adversely affects the overall efficiency of thin-film PV cells. If light absorption could be improved in thin-film cells of Si substrate, it could directly lead to higher energy conversion efficiencies.

II. THE EFFICACY OF NANOSTRUCTURES

To improve light-coupling efficiency, patterned nanostructures are designed and implemented in tandem with PV cells. These structures can be made of different materials depending on their optical properties, and are designed to enhance the light coupling efficiency. The nanoscale dimensions of these structures mean that they have a high surface-to-volume ratio, which complements their overall purpose. The usage of nanostructures provides a significantly cost-effective method to improve cell efficiency, and simultaneously improve efficiency beyond theoretical limits (Shockley- Queisser limit) [7].

Examples of nanostructured solar cells include

a) Nanowire solar cells, with nanopillars placed periodically on top of the Si substrate

b) Quantum dot solar cells

c) Plasmonic solar cells

d) Mesoscopic solar cells

The advantages and challenges of these designs are discussed in Figure 1 [4].

Nanostructures, in many cases, implement principles that are adhered to in other fields to enhance optical and electrical efficiency of solar cells. For example, moth-eye anti-reflective coatings (ARC) are inspired by the structure of the cornea of nocturnal moths which they utilize to have better vision in the dark [8] – [9]. Mesoscopic solar cells couple photo-sensitive pigment within the PV cell itself to provide a cheap thin-film solar cell that in recent studies have proved to be flexible in nature, allowing them to be coated on other machineries and equipment [10]. Organic solar cells utilize materials synthesized from organic substances or polymers that are optically and electronically active, facilitating them to be used as an environmentally friendly source of electrical energy [11].

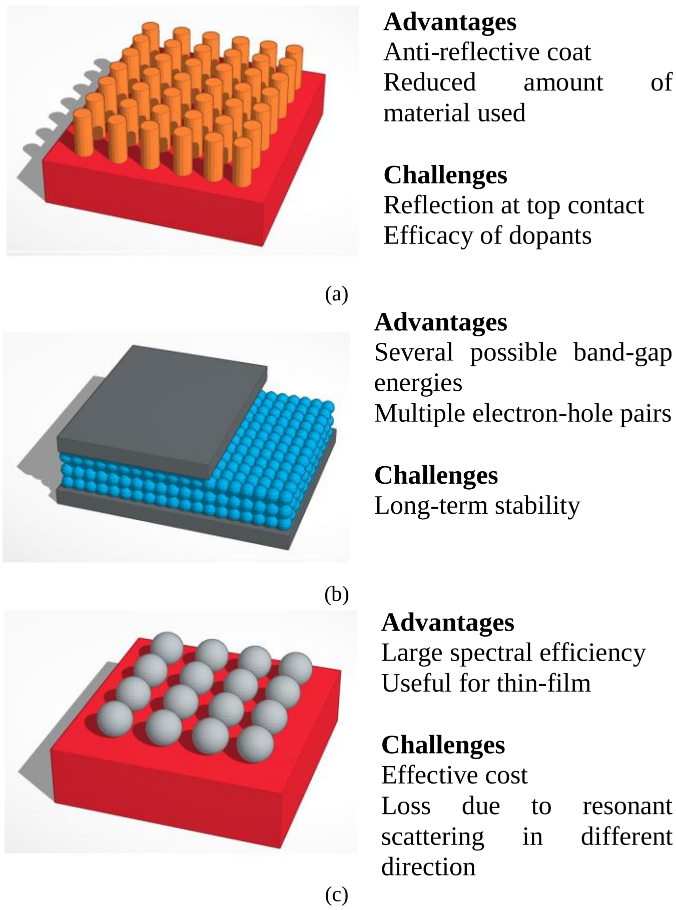


Figure 1. Examples of nanostructured solar cells (a) nanowire solar cells, (b) quantum dot solar cells, and (c) plasmonic solar cell

III. Plasmonic Solar Cells

The use of plasmonics in a variety of applications have led to significant research and development in this field over the past three decades. Surface plasmon polaritons (SPP) are generated by electromagnetic fields due to the oscillation of electrons on a metal-dielectric interface [12]. The use of these SPPs in improving PV efficiency has garnered great interest and research [13]. Studies conducted recently demonstrated the optical absorption and photocurrent generation in semiconductor photodiodes induced by scattering from SPP resonances in metallic nanoparticles deposited on the photodiode surface. Furthermore, they have been coupled with amorphous thin-film Si cells as well [14]. In order to properly optimize the usage of SPPs, detailed analysis of the dependence of the improvement of PV efficiency on the physical parameters of plasmonic nanostructures must be performed. This study will provide a brief analysis of this optimization as an example of how plasmonic nanostructures can improve the optical and electrical efficiencies of solar cells in general, and that this can be achieved by modifying physical parameters of the nanostructures used.

Plasmonic solar cells allow the thickness of the PV cell to be less than 2μm. Surface plasmon polaritons undergo resonance at certain frequencies, depending on the metal with which the nanostructure is made and the size and shape of the nanostructure [13]. At this state, SPPs demonstrate maximum resonance (e.g., scattering) with the incident electromagnetic waves (e.g., sunlight), which results in greater absorption within the cell substrate. This phenomenon can be manipulated to improve light coupling efficiency within thin-film solar cells. The metal with which the plasmonic nanoparticles are made is important, as different metals exhibit distinguishable surface plasmon resonances, and thus have varying impacts on the overall absorption enhancement. Below are examples of three possible configurations to incorporate plasmonic nanoparticles/nanostructures in thin-film Si solar cells:

a) Nanoparticles on the surface the substrate

b) Nanoparticles within the substrate

c) Nanostructures as/on back contacts

A. *Optimization of Parameters*

As part of this study, we wished to observe the overall influence that the physical parameters of plasmonic nanoparticles distributed periodically over a Si substrate had on the substrate's light coupling efficiency. In order to do so, we defined a quantity that would be an appropriate indicator for any form of enhancement or otherwise. Hence, we defined the parameter of absorption enhancement, g, which is the ratio of the power absorption within the Si substrate with the nanoparticles to that without the nanoparticles (g = power absorbed within the Si substrate with nanoparticles/power absorbed within the Si substrate with no nanoparticles). We then chose to calculate the short circuit current density, J_{SC}, to analyze how the absorption enhancement translates to the electrical energy conversion efficiency. Furthermore, to visually analyze the flow of the electric fields within the substrate as a result of the interaction of the plasmonic nanoparticles on the Si substrate, we generated and analyzed the near-field enhancement images.

B. Simulation Setup

All of the simulations were performed with cells placed under a plane wave source at a solar spectral irradiance of AM1.5G, with the average intensity taken to be 1000 W/m2 at 25°C. The analysis was done using the solvers FDTD Solutions (for optical absorption enhancement) and DEVICE (for the calculation of J_{SC}), developed by Lumerical Solutions, Inc. For each simulation, boundary conditions of anti-symmetric in the x-axis, and symmetric in the y-axis, were chosen to reduce simulation time. The simulation setup is illustrated in Figure 2.

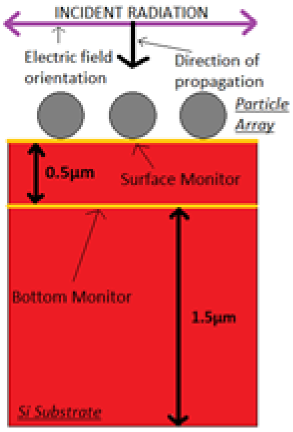


Figure 2. Simulation setup

C. Results and Discussion

Initially, we analyzed the absorption, scattering and extinction spectra for three metals (e.g., silver – Ag, gold – Au and aluminum – Al) with diameters of 50nm, 100nm and 200nm respectively to determine the exact range of wavelengths over which their resonances occur. The particles were then placed periodically over a 2µm thick Si substrate at a pitch (center-center) of 300nm. The absorption enhancement graph for the Si substrate coupled to each type of metal with the diameters of interest were then generated. The pitch was kept constant in order to observe the variation in optical absorption with respect to the diameter of the nanoparticles. The source wavelength range was chosen to be 400nm to 1100 nm for Ag and Au, and 200nm to 1100nm for Al as Al exhibited resonance on the bluer region of the spectrum. The results obtained are illustrated in Figure 3.

It can be seen that there is almost no absorption enhancement for Si substrate with nanoparticles of diameter of 50nm on the surface for all three metals. The absorption enhancement factor is observed to have a peak for 100 nm Ag at approximately λ ~ 420nm and for 200 nm Ag at approximately λ ~ 470nm, while the same for Au is around 560nm (diameter = 100nm) and no clear peaks for the D = 200nm nanoparticle, respectively. For Al, the peaks lie closer to the λ ~ 350 nm. This is in agreement with our scattering spectra analysis (data not shown), as the plasmon resonance for each of these particles are around the same wavelengths.

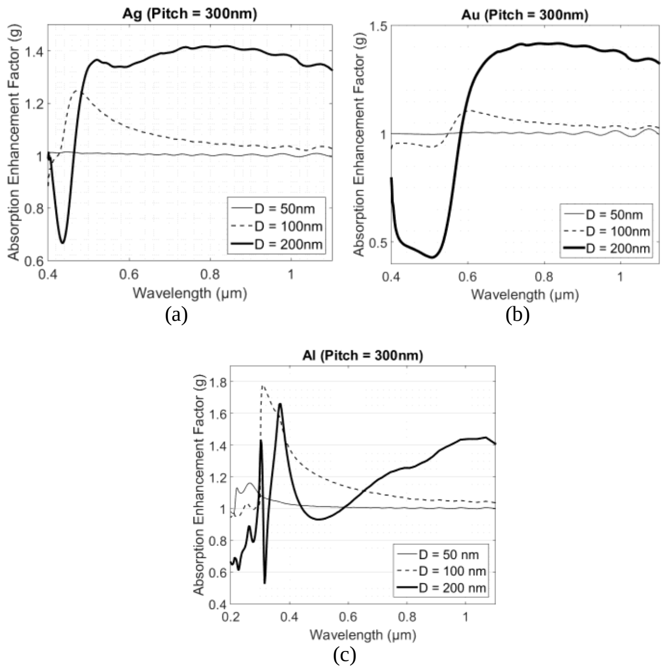


Figure 3. Absorption enhancement graphs for (a) Ag, (b) Au, and (c) Al at a pitch of 300nm

The analysis was then repeated for a pitch of 220nm (center-center). Absorption enhancement graphs were once again generated for all three metals with the three different diameters, and the results are shown in Figure 4. The absorption enhancement for particles of diameter 50nm is the same as before.

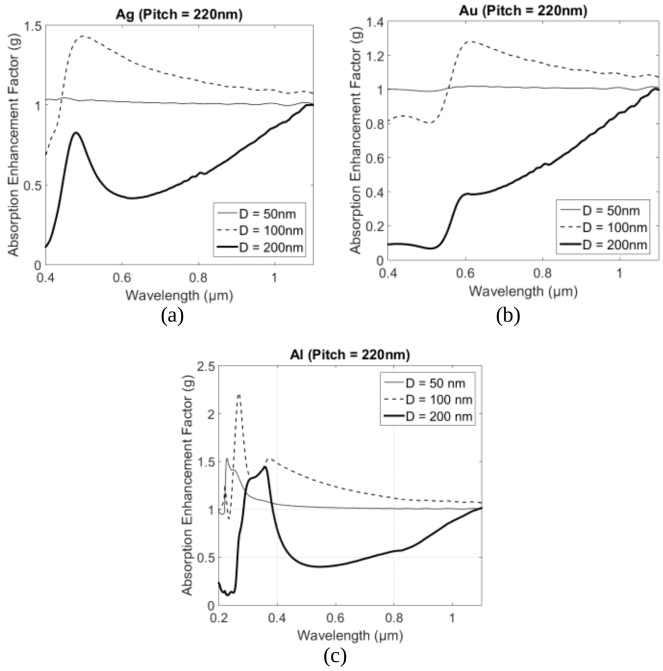


Figure 4. Absorption enhancement graphs for (a) Ag, (b) Au, and (c) Al at a pitch of 220nm

For particles of diameter 100nm, however, a greater enhancement is seen, as the peak absorption enhancement factors for all three particles are now greater than it was with a pitch of 300nm. Furthermore, it is observed that for particles with diameter 200nm, the absorption enhancement factor is less than 1 for a greater part of the spectrum. One possible explanation for this could be that for this diameter of metallic nanoparticles, the scattering is occurring towards a direction away from the Si substrate.

The short circuit current density J_{SC} was calculated for all of the possible physical parameter combinations with the three different types of metals (Ag, Au and Al), three diameters (50nm, 100nm and 200nm) and two separate pitches (300nm and 220nm). J_{SC} was also calculated for the nanostructures at an incident wavelength range of 400nm to 700nm for Ag and Au, and 200nm to 600nm for Al, as it is within these ranges that the solar cell exhibited the highest respective absorption enhancements. The resultant J_{SC} are given in the following tables.

TABLE I. J_{SC} FOR THE LARGER INCIDENT SPECTRUM

(a)
For Pitch = 300nm

Structure	Ag (λ=400nm to 1100nm)		Au (λ=400nm to 1100nm)		Al (λ=200nm to 1100nm)	
	***J**SC*	***% Change***	***J**SC*	***% Change***	***J**SC*	***% Change***
No Particles	76.2793	0	76.2793	0	109.366	0
D = 50nm	78.2283	2.555	76.5712	0.383	111.636	2.076
D = 100nm	**95.5242**	**25.230**	**76.6562**	**0.494**	**135.808**	**24.178**
D = 200nm	41.1047	-46.11	19.2681	-77.362	59.7525	-45.365

(b)
For Pitch = 220nm

Structure	Ag (λ=400nm to 1100nm)		Au (λ=400nm to 1100nm)		Al (λ=200nm to 1100nm)	
	***J**SC*	***% Change***	***J**SC*	***% Change***	***J**SC*	***% Change***
No Particles	76.2793	0	76.2793	0	109.366	0
D = 50nm	77.3442	13.961	76.4398	0.210	110.743	1.259
D = 100nm	**89.7007**	**17.595**	**77.0183**	**0.969**	**127.488**	**16.570**
D = 200nm	77.0193	0.970	48.9927	-35.772	122.772	12.258

The most significant rise in J_{SC} is found to be for Ag nanoparticles with a diameter of 100nm placed at a pitch of 220nm, which is in compliance with previous studies that have been made. Furthermore, in conjunction with our observations for absorption enhancement, it is found that J_{SC} is reduced significantly for nanoparticles of diameter 200nm and pitch of 220nm. For the analysis with reduced spectrums of 400nm to 700nm for Ag and Au, the increase in J_{SC} is lower.

TABLE II. J_{SC} FOR LIMITED INCIDENT SPECTRUM

(a)
For Pitch = 300nm

Structure	Ag (λ=400nm to 700nm)		Au (λ=400nm to 700nm)		Al (λ=200nm to 600nm)	
	***J**SC*	***% Change***	***J**SC*	***% Change***	***J**SC*	***% Change***
No Particles	64.1602	0	64.1602	0	60.0142	0
D = 50nm	65.1141	1.487	63.8908	-0.420	61.1944	1.967
D = 100nm	**75.6064**	**17.840**	**59.8337**	**-6.743**	**75.1950**	**25.295**
D = 200nm	59.2331	-7.679	26.0504	-59.398	61.7457	2.885

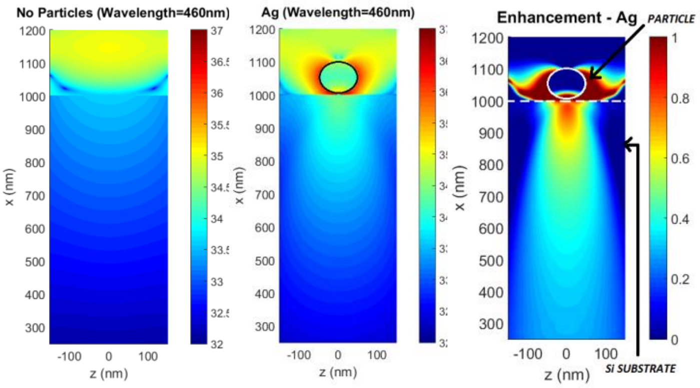
(b)
For Pitch = 220nm

Structure	Ag (λ=400nm to 700nm)		Au (λ=400nm to 700nm)		Al (λ=200nm to 600nm)	
	***J**SC*	***% Change***	***J**SC*	***% Change***	***J**SC*	***% Change***
No Particles	64.1602	0	64.1602	0	60.0142	0
D = 50nm	65.8854	2.689	63.6557	-0.786	62.1060	3.486
D = 100nm	**79.7938**	**19.691**	**56.1374**	**-12.504**	**81.4004**	**35.635**
D = 200nm	32.9415	-48.657	8.44858	-86.832	32.2894	32.389

This can be attributed to the fact that the absorption enhancement is occurring across all the wavelengths (frequencies) studied, and thus curtailing the frequency range also lowers the percentage increase in J_{SC} for the case of Ag and Au. For Al, the opposite phenomenon occurs. This may be explained by the fact that the majority of the absorption enhancement for Al occurs in the limited incident spectrum range (i.e., λ=200nm to 600nm range).

D. Near-Field Enhancement

Near-field plots would demonstrate the transmission of the electric field within the nanoparticle and the Si substrate at the resonant wavelength of the particle. To this end, we chose the physical parameters at which each Ag nanoparticle exhibited the highest absorption and J_{SC} enhancement. This was found for Ag nanoparticles with a diameter of 100nm placed at a pitch of 220nm. We also had to first identify the approximate wavelength at which the nanoparticles of Ag displayed surface plasmon resonance ~ λ = 460nm. Three near-field images were generated – (a) the substrate without the particle at the given wavelength (b) the substrate with the Ag particle at the given wavelength and; (c) the enhancement image, which was calculated by dividing the raw data of the second image (Fig. 5b) with the raw data of the first image (Fig. 5a). All three images in that exact order are shown in Figure 5. The images were generated at a wavelength of λ = 460nm, which is approximately the wavelength at which the plasmon resonance phenomenon for Ag nanoparticle of diameter 100nm occurred. For the enhancement image, the color scale is in the log scale, and hence the areas which are dark red color have an enhancement of 1 in the log scale that corresponds to a near-field enhancement of x10 (10 times).



(a) (b) (c)

Figure 5. Near-field plots for (a) without particles, (b) with Ag particles of diameter 100nm, and (c) enhancement image, at λ=460nm

The near-field enhancement image (Fig. 5c) shows us the presence of the Ag nanoparticle on top of the Si substrate leads to significant increases in the electromagnetic field intensity around the Ag nanoparticle and within the Si substrate in the vicinity of the Ag nanoparticle. It is important to note that the near-fields calculated in Figure 5(b-c) can represent enhanced propagating radiation within the Si substrate. In Figure 5c, a significantly large portion of the image around the Ag nanoparticle is bright or dark red which depicts significant near-field enhancements in the region immediately surrounding the Ag nanoparticle which is induced by the incident radiation. These near-field enhancements which also carry into the Si substrate immediately surrounding the Ag nanoparticle can eventually lead to the absorption enhancements and increase in J_{SC} as shown in Figures 3 and 4, and Tables I and II. Hence we see that nanoparticle systems displaying large absorption enhancements and increase in J_{SC} also show very strong enhancements in the near-fields around the Ag nanoparticle.

IV. COLNCLUSION

This study provides a brief review of the different methods of using patterned nanostructures to improve light-coupling efficiency in PV cells (e.g., thin film Si solar cells). The patterned nanostructures described can be designed and implemented in tandem with PV cells. These structures can be made of different materials depending on their optical properties, and are designed to enhance the light coupling efficiency of the PV cell. The different kinds of patterned nanostructures discussed include nanowire solar cells with nanopillars placed periodically on top of the Si substrate, quantum dot solar cells, mesoscopic solar cells, and plasmonic solar cells. The review discussed the advantages and disadvantages of using the different kinds of patterned nanostructures for enhancing PV cell efficiency.

In particular, this study expands upon the use of plasmonic metal nanoparticles for increasing the efficiency of thin-film Si solar cells [15]. This study explores the relationship between energy conversion efficiency in amorphous thin-film Si solar cells and the type of metallic nanoparticle, the size of the metallic nanoparticles, and the distance between neighboring metallic nanoparticles in a periodic array of the metallic nanoparticles deposited on top of the Si substrate. The different configurations of metallic nanoparticle systems studied were as follows: D = 50 nm, D = 100 nm, and D = 200 nm Ag nanoparticles spaced with a pitch of 300 nm (center-center) and a pitch of 220 nm (center-center); D = 50 nm, D = 100 nm, and D = 200 nm Au nanoparticles spaced with a pitch of 300 nm (center-center) and a pitch of 220 nm (center-center) and; D = 50 nm, D = 100 nm, and D = 200 nm Al nanoparticles spaced with a pitch of 300 nm (center-center) and a pitch of 220 nm (center-center), respectively. Of the different metallic nanoparticle systems studied, it was found that Ag nanoparticles of D = 100 nm and spaced 220 nm (center-center) from neighboring nanoparticles in a periodic array of the metallic nanoparticles deposited on top of the Si substrate provides the maximum increase (i.e., 25%) in the generated short circuit current density (J_{SC}) in a specific wavelength region of the electromagnetic spectrum (λ = 400 – 1100 nm). It was also shown that Al nanoparticles of D = 100 nm and spaced 220 nm (center-center) from neighboring nanoparticles in a periodic array of the metallic nanoparticles deposited on top of the Si substrate provides the maximum increase (i.e., 35%) in the generated short circuit current density (J_{SC}) for the bluer wavelengths of the electromagnetic spectrum (i.e., of λ = 200 – 600 nm). This can be attributed to the plasmon resonance wavelength of Ag (λ ~ 430 nm) being more red-shifted than that of Al (λ ~ 320 nm). So the increase in J_{SC} appears to be strongly correlated to the plasmon resonance wavelength of the metal nanoparticles. Additionally, this study also shows the near-field enhancement image for a D = 100 nm Ag nanoparticle on top of a Si substrate (Figure 5a-5c). The near-field enhancement image (Figure 5c) shows the presence of the Ag nanoparticle on top of the Si substrate leads to significant increases in the electromagnetic field intensity around the Ag nanoparticle and within the Si substrate in the vicinity of the Ag nanoparticle when the system is illuminated by incident electromagnetic radiation. This observation suggests that such near-field enhancements can eventually lead to the absorption enhancements and increase in J_{SC} as shown in this study.

ACKNOWLEDGMENT

The authors would like to sincerely thank Dr. Khosru M. Salim of the Dept. of Electrical and Electronic Engineering of Independent University, Bangladesh (IUB) for lively discussions and funding for S.A.C. The authors would also like to thank IUB for providing assistance and funding for this research.

REFERENCES

[1] S. Pillai, and M. A. Green, "Plasmonics for photovoltaic aplications," Solar Energy Materials & Solar Cells, vol. 94, pp. 1481-1486, 2010.

[2] J. Nelson, The Physics of Solar Cells. London, England: Imperial College Press, 2003.

[3] P. Wüffel, Physics of Solar Cells:, Principles to New Concepts. New York, NY: Wiley-VCH, 2004.

[4] M. C. Beard, J. M. Luther and A. J. Nozik, "The promise and challenge of nanostructured solar cells," Nature Nanotechnology, vol. 5, pp. 951-954, 2014.

[5] T. Tiedje, E. Yablonovitch, G. D. Cody, and B. G. Brooks, "Limiting efficiency of silicon solar cells," IEEE Transactions on Electron Devices, vol. 31, issue 5, pp. 711-716, 2005.

[6] A. G. Aberley, "Thin-Film olar Cells," Thin Solid Films, vol 517, issue 17, pp. 4706-4710, 2009.

[7] W. Shockley, and H. J. Queisser, "Detailed Balance Limit of Efficiency of p‐n Junction Solar Cells," Journal of Applied Physics, vol. 32, issue 510, 1961.

[8] C. H. Sun, P. Jiang, and B. Jiang, "Broadband moth-eye antireflection coatings on silicon," Applied Physics Letters, vol. 92, 2008.

[9] S. A. Boden, and D. M. Bagnall, "Optimization of moth-eye antireflection schemes for silicon solar cells," Prog. Photovolt: Res. Appl., vol. 18, pp. 195–203.

[10] N. G. Park, "Organometal Perovskite Light Absorbers Toward a 20% Efficiency Low-Cost Solid-State Mesoscopic Solar Cell," J. Phys. Chem Letters, vol. 4, pp. 2423–2429.

[11] Y. J. Cheng, S. H. Yang, and C. S. Hsu, "Synthesis of Conjugated Polymers for Organic Solar Cell Applications," Chem. Rev., issue 109 (11), pp. 5868–5923.

[12] W. L. Barnes, A. Dereux, and T. W. Ebbesen, "Surface plasmon subwavelength optics," Nature, vol. 424, pp. 824-830.

[13] S. Pillai, K. R. Catchpole, T. Trupke, and M. A. Green, "Surface plasmon enhanced silicon solar cells," J. Appl. Phys., vol. 101, 093105, 2007

[14] Y. Wang, T. Sun, T. Paudel, Y. Zhang, Z. Ren, and K. Kempa, "Metamaterial-Plasmonic Absorber Structure for High Efficiency Amorphous Silicon Solar Cells," Nano Lett., vol. 12 (1), pp 440–445, 2012.

[15] P. Spinelli, V. E. Ferry, J. van de Groep, M. van Lare1, M. A. Verschuuren, R. E. I. Schropp, H. A. Atwater and, A. Polman, "Plasmonic light trapping in thin-film Si solar cells," J. of Optics, vol. 14 (2), 2012.

WoTCoMS: A Novel Cross-Layered Web-of-Things Based Framework for Course Management System

Nusrat Jahan Farin[1], Atiqur Rahman[2], Nafees Mansoor[3], Sazzad Hossain[4]
[1, 3, 4] Department of Computer Science and Engineering, University of Liberal Arts Bangladesh
[2] School of Engineering & Computer Science, Chittagong Independent University
E-mail: [1]nusrat.farin.cse@ulab.edu.bd, [2] arahman@ciu.edu.bd, [3]nafees.mansoor@ulab.edu.bd, [4]sazzad.hossain@ulab.edu.bd

Abstract— **A course management system (CMS) is a tool that facilitates faculties and students of any educational institution to develop and to support online education management system. By creating a virtual classroom, CMS provides an online platform where students and faculties can share and can access course resources. Moreover, functionalities that come with managing a large class room can be organized and controlled by using such system. Due to its flexibility, interactive features and delivery medium satisfaction, CMS is getting more popularity over the years. Thus, incorporating CMS with Web of Things (WoT), the cutting edge web technology, is getting more importance in the research community. This paper proposes a novel cross-layered WoT based CMS framework, named "WoTCoMS". Design and operational features of the proposed WoTCoMS have been highlighted in this paper. While designing the proposed framework, different layers of WoT are considered, however, the proposed system is designed as a cross-layered approach. The proposed CMS also considers that strengths and limitations of the existing CMSs. To the best of our knowledge, this is the first course management system with the integration of the features of WoT.**

Keywords—Course Management System (CMS); Web of Things (WoT); Web-based learning; WoTCoMS; E-learning

I. INTRODUCTION

Now-a-days web of things (WoT) has become one of the most widely used areas for research. The terms web of things is used to define that any object of the world is allowed to use/ access the World Wide Web (www). Web of things is a layer based model where the layers are impactful in real world. In general there exist five layers in WoT, Where Layer 0 contains the networked things like teleconferencing, RFID code etc. Layer 1 which support to access coding part like PHP, Android, Windows etc. Layer 2 actually contains the finding parts like Rest crawler, Semantic web, Search engine and so on. Sharing part of a system such as social network, authentication, encryption etc. is contained by layer 3. Layer 4 contains the composing parts namely web application, system integration etc. The WoT is very important for online course management system.

The utilization of online instruction frameworks has increased exponentially in the last couple of years. Online course management system (CMS) is a tool or engine that allows instructors, universities, and corporations to develop and to support online learning or education. Specially, shared and specialized tools of CMS are generally utilized as a part of instructive settings. Therefore, Virtual Learning Environments (VLE) are introduced more by universities, community colleges, schools, businesses, and even individual educators. In some cases, such e-learning frameworks are called as Learning Management System (LMS), Managed Learning Environment (MLE), Learning Support System (LSS), Course Management System (CMS), Learning Platform [1] or Learning Content Management System (LCMS) and so on [7].

E-learning is much more effective to maintain all the tasks properly of a classroom based education system. There are many factors in a bigger classroom such as large number of enrolled students, role type of various faculty members, different types of assistantships, grades, assignments, class lectures, assignment collection, scores etc. Learning management system accumulates an extraordinary arrangement of log information about understudies' exercises [5]. Many institutions choose to use e-learning because of many reasons such as - (1) providing consistent and worldwide training, [2] lessen conveyance process duration, (3) building learners' comfort, (4) reducing cost and so forth.

Thus, considering the benefits of WoT, a framework for course management system named "WoTCoMS" is proposed in this paper. Though, there are several different course management systems, however all of these CMSs fail to extract the maximum benefit of WoT technology in their design. For instance, in many CMSs, there is no feature to see all course results at a glance. Moreover, existing CMSs are also incapable to notify run time updates to the users in different social media platforms or/and incapable to notify the users using the existing telecommunication systems. Therefore, this paper proposes a novel WoT-based framework for course management system. This proposed system includes the standards for identification, discovery and interoperation of the services across platforms from several vendors. It also involves the need for rich descriptions and shared data models, as well as concern about the security, privacy, scalability, accessibility and many things. Different layers of web of things are considered while designing the proposed WoTCoMS system. Moreover, while designing the proposed framework, these layers are merged to make the system more efficient. Thus, WoTCoMS is considered as a cross-layered WoT-based course management system. The layered diagram of this proposed system is shown in Figure 3. The Design aspects and features of the proposed WoTCoMS are discussed thoroughly in the later section of this paper.

The paper is organized as follows. In section II, a brief study on the existing CMS tools has been presented. This section also discusses about different CMS tools with their pros and cons. Section II also highlights current global trends of CMS usages. The existing state of CMS usages in Bangladeshi University is highlighted in section III. The proposed model is presented in Section IV. Conclusion and future works have been discussed in section V.

II. Background Study

Currently different types of Learning Management Systems (LMS) [1] or Course Management Systems (CMS) are used such as Moodle [2], Blackboard [6], Edmodo [3], Schoology, etc. Moodle is much more popular than the other CMS. Nearly 74% of associations currently use Learning management systems (LMS) and Virtual classroom or web-casting or video broadcasting or course management system (CMS) [2].

Table1. The growth rate of CMS usage country

Rank	Country	Growth rate (Percent)
1	India	55
2	China	52
3	Malaysia	48
4	Romania	38
5	Poland	28
6	Czech Republic	27
7	Brazil	26
8	Colombia	25
9	Indonesia	20
10	Ukraine	20

Online learning management places the most important role for many academic training or courses, where many CMS are used for the academic purposes, which contain many features. Most of the CMS have user friendly interface, personalized dashboard, convenient file management system, simple and intuitive text editor, notifications or get alerting features, tracking progress, customizable site design and layout, secure authentications, multilingual capability, roles & permission, direct learning path, encourage collaboration, group management, marking workflow, assignment module, workshop, badges, grade booking, embedded external resources and so on. All of these features are the benefits of these CMS. There are also many drawbacks in existing CMS. It is worth to mention here, the admin side of the CMS is not much user friendly. Thus, it is not so easy to maintain all the features properly. The existing CMS have no easy way to manage groups of students whereas it would be easier if there exist a way to manage groups site wide rather than course wise. LMS has no grading system that can show all grades as a report to the user. User can only see the grade of each of the courses at a time. Current researches on e-learning or course management systems focus e-learning management system [7], design of e-learning [10], evolution of e-learning [11], impact of online learning management system [17-18], online management to improve classroom [17], characteristic of usage e-learning [11], [20], cloud computing based e-learning [13-14], web based learning management system [15-16],[18], [21-22] and so on.

E-learning can be characterized as the utilization of PC system innovation, essentially over an intranet or through the Internet, to convey data and direction to people. To highlight the importance and its trends of CMS development rate of self-managed e-Learning by developing nation is given with a Table **1** [4]. As a test case, Bangladesh is taken into consideration, where current state of CMS is highlighted in the next section.

III. Existing State of CMS Usages in Bangladeshi University

Course management system is used by some of the private and public universities in Bangladesh. Developing countries all over the world uses the course management system very in a wide scale. Comparing with other countries, Bangladesh has some lacking to use the online course management system. Some private and public universities use the CMS but the number of the users is not that high and many of the universities use CMS partially. In Fig. 1, all universities of Bangladesh are denoted with a pie chart. Our survey shows that almost 70% of the universities don't use CMS which is denoted with red in Fig. 1. Few departments of some universities use CMS partially, which is denoted as partially in the figure. Approximately 9% of the universities are using CMS. Only 21% universities use the CMS properly and regularly as shown in Fig. 1 with blue zone.

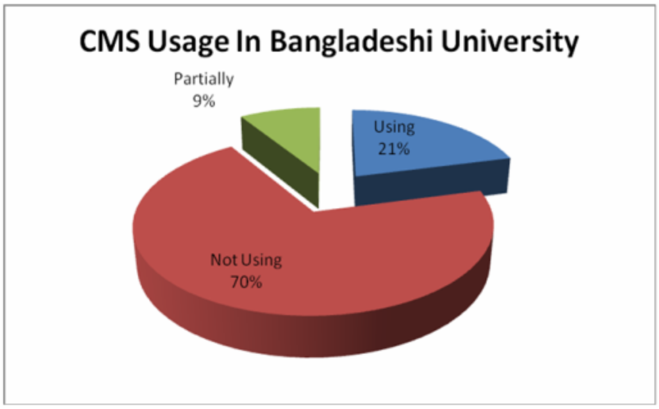


Fig. 1 Current state of CMS usage in Bangladeshi universities

In Bangladesh the universities use different platforms of CMS. Fig. **2** shows the usage ratio of the different platforms users of the course management system. Most of the CMS users use moodle [1-2], [22] in Bangladesh. Black box and Schoology are also used. Other types of platform are also used such as edmodo [3], google classroom etc. However usage of these platforms is quite rare among all of the existing course management systems. As shown in Fig. 2, it is observed Moodle is the most popular in Bangladesh. After analyzing the strengths & limitations of existing CMS, WoTCoMS is proposed and presented in this paper. The proposed system also considers standards for identification and interoperation of the services over the different platforms.

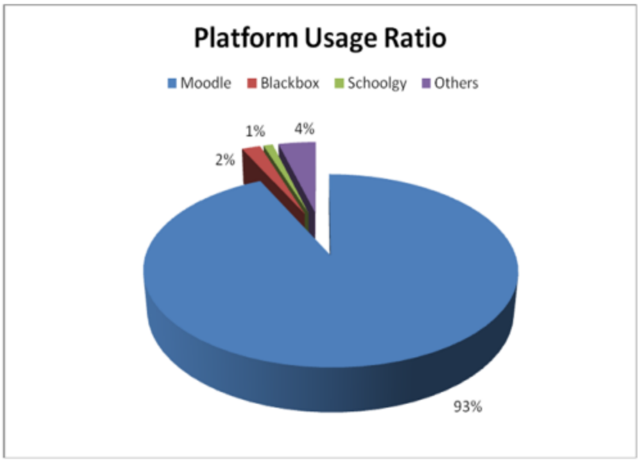


Fig. 2 Usage ratio of different CMS Platforms in Bangladesh

WoTCoMS also involves the need for flourishing descriptions and shared data models, as well as concern about the authentication, privacy, scalability, accessibility. The proposed WoTCoMS is described in the following section.

IV. Proposed Framework

This section is discussed about the model that is proposed in this paper. The overview of the proposed framework is detailed with a flowchart view in Fig. 4. The key features of the proposed CMS tool is briefly discussed here.

A. *Portability*

Most of the web-based CMS have the flexibility to run on any platform like Linux, Windows, Mac OSX and so on. The proposed system is design with the help of existing system and also adds some other features to ensure portability.

1) *Platform*

Considering platform dependency, the proposed system is designed as platform independent. Thus, it can run on any OS such as windows, Linux, Mac OSX etc.

2) *Environment*

This system works in different environments like iOS application, android apps, windows app, desktop app and web application.

B. *Design Features*

While designing the proposed system, firstly we are interested to understand how the system will be used and how the proposed system will be able to fulfill the requirements of the users. Thus, some policies and rules are integrated in this system that can override on case-by-case basis and also can be able to notify the users (particular user) of the updated information via mobile phone and social media.

1) *Teleconferencing*

This option can be used by both teacher and students. By using this feature, the users are able to make a teleconference or video call. This features is quite unique as it has one to one conferncing and many to many conferncing abilities. Our present research shows very few systems has this option.

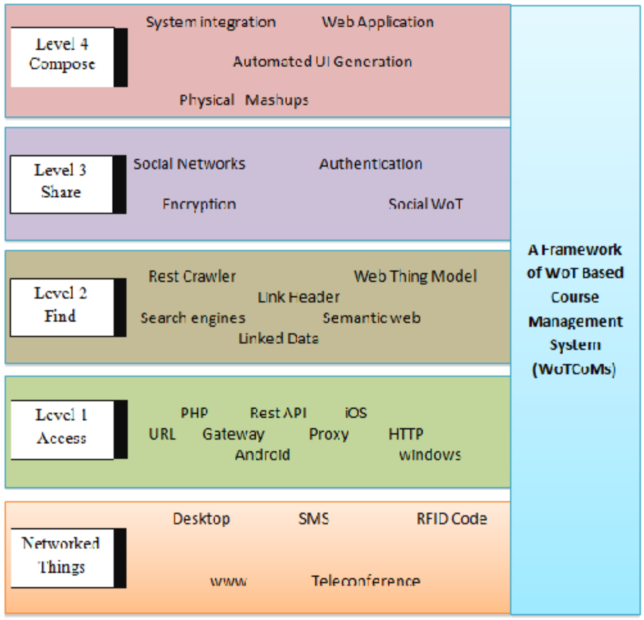


Fig. 3 Cross-layered Design of the proposed Model

2) *Plagarism Check*

This option gives the opportunity to the faculties to check plagiarism of the individual or group assignments. It is a very important feature for any course management system.

3) *Search Engine*

In this section faculty, admin and students can search any keywords for this system. By using this section user can optimize their search engine.

4) *Crawler*

By using this section sudents and faculties are able to crawl among several libraries such as world research library, university's online library, some other digital library and so on.

5) *Dsicussion Forum*

This option gives the opportunity to make discussion forum for the students. Users can give their philosophy or view and share with the entire user by using this feature.

6) *User friendly interface*

It is very important to have a system with user friendly graphical view that can help the users to use the system easily to learn and also can provide the environment to work perfectly in a secure way. The proposed system is designed with a very user friendly graphical interface.

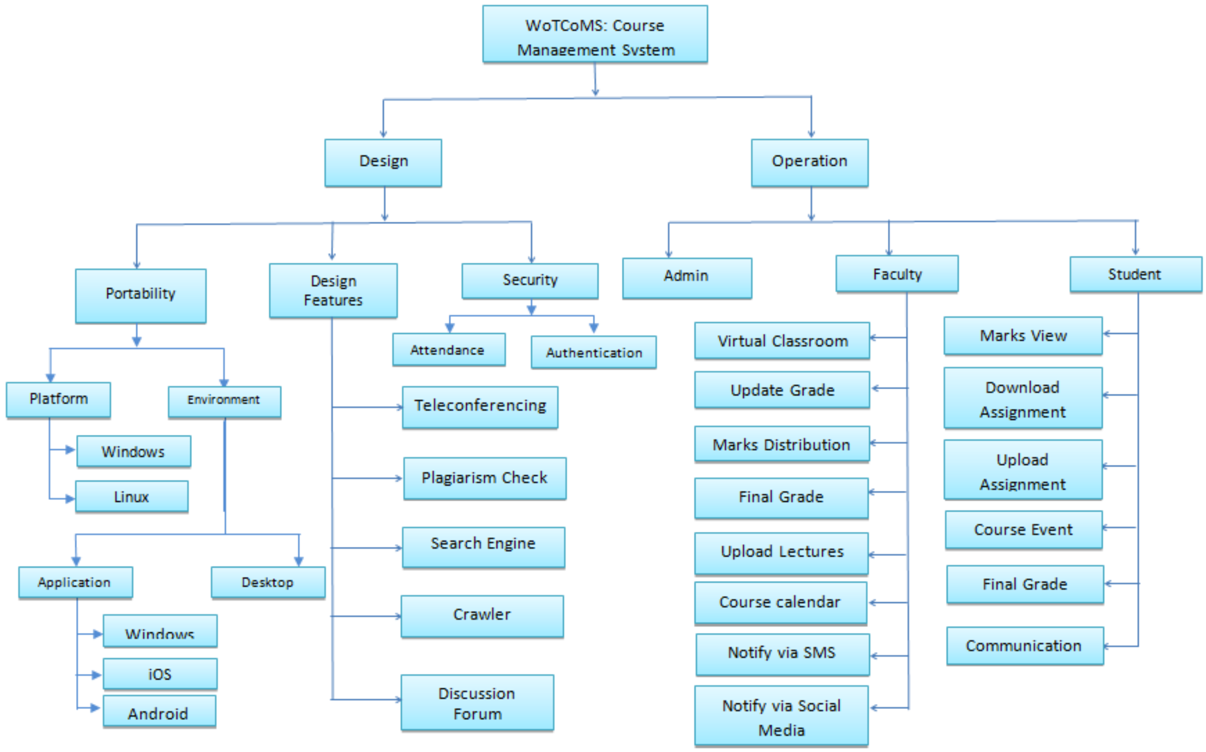


Fig. 4 Overview of the proposed CMS

C. *Security*

1) *Authentication*

"WoTCoMS" users are authenticated by their identification number. Fine-grained access control model also support this system. It is design with several level of access for the course staff, students and the admin officer. Any entity, which is not authorized, cannot access the course material.

2) *Attendance*

This system can track the student RFID. In WoTCoMs the attendance of the student can be taken by tracking the RFID of the student.

3) *Course management for the administrator and the user*

The administrator is responsible for adding, deleting or updating any information in the proposed system. For updating any course info, students or faculties information, or any kind of changes the administrator must be authenticated and after that they can able to do the changes. At the end of the each semester the faculties are able to update the grade of the students and the system can automatically make a final mark sheet in a file which contains all courses grade for each of the students. At the end of the semester the students whose registration date is finished are not able to access the particular courses. However they can check their grade sheet or others information.

D. *Course management system for the faculties view point*

Faculty memebers must need to register to use this system. After completing their registration the memebers are able to select the courses and their consultation hour from the system. After finishing all these steps, faculty can manipulate or maintain the CMS by doing the following steps.

1) *Virtual classroom management*

This section is used for placing the marks distribution for a particular course by a faculty member. When the faculty member enters into this section s/he can see the registered students for this particular course. And only the registered students can access the course material. Faculty can add some students if s/he wish to add unregistered students to this course. This features gives extra flexibity for the faculties. From this section the teacher can also take attendance in online rather than the manual one. There is also another feasibility, where once the students open their RFID the students automatically detect in this system and assign as an attendee.

2) *Updating grade system*

Grading policy may vary from terms to terms, course to course and faculty to faculty. This section gives the opportunity to the faculty members to set the grading system in their own way for a particular subject or course.

3) *Marks distribution policy*

Sometimes it is needed to change the grading policy and this section gives the user to change their grading policy or can add new content or new policy.

4) *Final grade calculation*

In this section faculty can overview the grade of all students in a file. If any faculty wants to change any marks of the students they can change that particular part by using this option.

5) *Students marks*

To search any examination marks of a particular course or student of a particular exam, assignment or quiz, this section can be used.

6) *Uploading lectures, materials and assignments*

In CMS faculties need to upload their lectures, materials, assignments and so on for the students. This option can be used to upload their necessary things. The uploaded documents by the students like assignments, quiz paper can be downloaded by the faculties.

7) *Course calender*

Course calendar is one of the most important things for the online course management system. The faculty member can add the event details for the particular courses by using this section.

8) *Notify students with message*

By using this option faculty can be able to notify the students via text message using web to SMS technology.

9) *Notify students via social media*

If any faculty post the update anything to the CMS and wants to notify the students, faculty can be able to notify the students via social media like facebook, linkedIn, yahoo or gmail by using this option.

E. *Course management system for the students view point*

The particular students who are approved their registration by the administrations can access the proposed CMS and their particular courses. Each student requires ID number and a password to login to this system.

After login to the system, the students will find all the facilities of the online course management system. This system may create a virtual classroom for the students and the faculties. The students can enjoy the following facilities by using the proposed system.

1) *Marks view*

Students can see all their examination, quiz or assignments marks that are given by their course teacher. Students can also be able to find their grade on the basis of marks where the grade is automatically calculated by the system from the policy which is assigned by the course teacher.

2) *Downloading Assignments & Lecture Slides*

Students can download their necessary lectures, course materials or others thing by using this option.

3) *Lectures viewing*

Lecturers can record their lecture and upload to the system. Thus, using this option, students of a particular course can view any missed lecture or important lecture for better understanding the topic.

4) *Uploading assignment*

This section enables the students to upload their assignments for the courses.

5) *Course events*

Course calendar option gives the opportunity to the students to be informed about the upcoming course events. It may also work as notice board for the students.

6) *Final Grade*

At the end of the semester students can able to view their final grade.

7) *Communicate with the faculties*

By using this option the students can communicate with their course teacher by the chat box.

F. *Updating Notification*

This option is used by the faculty and also by the students to make the notification for any change or update in any things such as new upload lectures, assignments etc.

G. *Mobile apps*

This system can be also used by the mobile application. For any update, users can check their update through mobile application. Students can also submit their assignments using mobile app. Faculties can also upload their lectures using the mobile application to the proposed course management system.

V. CONCLUSION AND FUTURE WORKS

In this paper, a novel Cross-Layered Web-of-Things Based Framework for Course Management System is presented. The proposed model aims to provide all the features described in the earlier sections. Though there exist many CMS however features considered in the proposed CMS differs from others because of its uniqueness. It is worth to mention that features of the proposed “WoTCoMS” are very important for the CMS users and make the system very friendly and easier to use. Our next research step is to introduce WoTCoMS to different universities and based on the feedbacks enhance the proposed system. Moreover, more components of WoT are intended to be integrated with the proposed model.

REFERENCES

[1] Brian Beatty and Connie Ulasewicz. Faculty perspectives on movingfrom blackboard to the moodle learning management system. TechTrends, 50(4):36–45, 2006.

[2] Jason Cole and Helen Foster. Using Moodle: Teaching with the popular open source course management system. "O’Reilly Media, Inc.", 2007.

[3] Chada Kongchan. How a non-digital native teacher makes use of edmodo. In 5th Intenational Conference ICT for language learning, 2008.

[4] Ingo Lütkebohle. The Worldwide Market for Self-paced eLearning Products and Services: 2010-2015 Forecast and Analysis.

[5] Florence Martin. Blackboard as the learning management system of a computer literacy course. Journal of Online Learning and Teaching, 4[2]:138–145, 2008.
[6] Cristóbal Romero, Sebastián Ventura, and Enrique García. Data mining in course management systems: Moodle case study and tutorial. Computers & Education, 51(1):368–384, 2008.
[7] Kimball, Andrew. "E-learning system." U.S. Patent Application 11/370,297, filed March 7, 2006.
[8] Chen, C. M., Lee, H. M., & Chen, Y. H. (2005). Personalized e-learning system using item response theory. *Computers & Education*, *44*(3), 237-255.
[9] Ismail, J. (2001). The design of an e-learning system: Beyond the hype. *The internet and higher education*, *4*(3), 329-336.
[10] Shee, D. Y., & Wang, Y. S. (2008). Multi-criteria evaluation of the web-based e-learning system: A methodology based on learner satisfaction and its applications. Computers & Education, 50(3),894905.
[11] Liaw, S. S. (2008). Investigating students' perceived satisfaction, behavioral intention, and effectiveness of e-learning: A case study of the Blackboard system. Computers & Education, 51[2], 864-873.
[12] Özpolat, E., & Akar, G. B. (2009). Automatic detection of learning styles for an e-learning system. Computers & Education, 53[2], 355-367.
[13] Masud, M. A. H., & Huang, X. (2012). An e-learning system architecture based on cloud computing. system, 10(11).
[14] Dickey, M. (2004). The impact of web-logs (blogs) on student perceptions of isolation and alienation in a web-based distance-learning environment. Open learning, 19(3), 279-291.
[15] Shum, Simon Buckingham, and Rebecca Ferguson. "Social Learning Analytics." Educational technology & society 15.3 (2012): 3-26.
[16] Perkins, M., & Pfaffman, J. (2006). Using a course management system to improve classroom communication. SCIENCE TEACHER-WASHINGTON-, 73(7), 33.
[17] West, R. E., Waddoups, G., & Graham, C. R. (2007). Understanding the experiences of instructors as they adopt a course management system. Educational Technology Research and Development, 55(1), 1-26.
[18] McGill, T. J., & Klobas, J. E. (2009). A task–technology fit view of learning management system impact. Computers & Education, 52[2], 496-508.
[19] Machluf, Y., Gelbart, H., Ben-Dor, S., & Yarden, A. (2016). Making authentic science accessible—the benefits and challenges of integrating bioinformatics into a high-school science curriculum. Briefings in Bioinformatics, bbv113.
[20] Palumbo, E., & Verga, F. (2015). Creation of an integrated environment to supply e-learning platforms with Office Automation features. Interactive Learning Environments, 23(6), 766-777.
[21] Braga, F. L., Rodrigues, R., & Bolzan, M. S. (2015). Moodle and Physics learnig: A good experience with High School students. Lat. Am. J. Phys. Educ. Vol, 9(3), 3403-1.
[22] Mansoor, N., Rajib, F. U., Chowdhury, S. A., Al Mamoon, I., & Mansoor, N. (2008, June). MY SB: A Web-based Course Management System. International Conference on Electronics, Computer and Communication (ICECC 2008).

Simulation Based Study to Present the Performance of Ad-hoc Routing Protocol

Shad Muhammad, Ghani ur Rehman, Tariq Usman, Fawad Taj
Computer Science Department
Khushal Khan Khattak University, Pakistan

Atiqur Rahman
School of Engineering and Computer Science,
Chittagong Independent University, Chittagong,
Bangladesh.

Asim Zeb
Computer Science Department
Qurtaba University of Science and Information Technology, Peshawar, Pakistan

Abstract--Ad-hoc mobile/802.11 networks are considered as networks with no fixed physical line connections. Having no fixed topology and open to movement of the end nodes, Intervention, multiple paths creating and losing paths. Each node taking part in the network acts host and routing node and mobile nodes forward packets to other nodes. Thus some kind of routing protocols are needed for these mobile nodes to fully operate and function properly. Ad-hoc network requires routing protocol to send data to destination node. The significant one is the dynamic routing protocol, which quickly change the topology. The dynamic routing protocol minimizes the periodic update messages. Each node updates the routing table to send data to destination node. Reactive protocol searches a route to destination/remote node on needed basis while Proactive protocols maintain the whole routing table at each node. This study investigates the performance of two widely known ad-hoc routing protocols, namely AODV and DSR, in terms of packet delivery ratio, average end-to-end delay and routing overhead by changing the mobility. The simulation has been carried out using NS2 2.29 as the simulation platform.

Keywords—routing; Ad-hoc; mobile nodes; dynamic routing;

I. INTRODUCTION

Mobile ad-hoc network/802.11 has different applications, which can be used in commercial and industrial site. Some of the applications are defined below [2]. Important applications of ad-hoc applications are emergency services, commercial services, Education services, enterprise application, and Entertainment services.

Ad-hoc networks are implemented with type of remote data transmission system that uses some form of waves as a media, which are electromagnetic_and radio waves, for the carrier and this implementation normally takes place at the physical layer. In the last few years, the world networks have increasingly become a mobile. This is because the recent advancement in nodes such as laptops and PDA (personal data assistant), which has brought these nodes to the lower prices and increase the high data rates.

Ad-hoc networks can be characterized into two forms

I. Infrastructure network
II. Ad-hoc network.

In infrastructure mobile network, mobile nodes have wired base stations in a specific range. The base station contains the central controller for an infrastructure network. In contrast, mobile ad-hoc networks are self-organized networks without infrastructure support. Nodes move in a random manner, therefore the network may experience a quick and unknown topology changes. Furthermore, because nodes in a mobile ad-hoc network normally have limited communication range, some nodes will not send or receive data packets directly. Hence, routing paths in ad-hoc networks contain multiple hops, and every node in ad-hoc networks has the responsibility to act as a router and send and receive the data packets. [1]

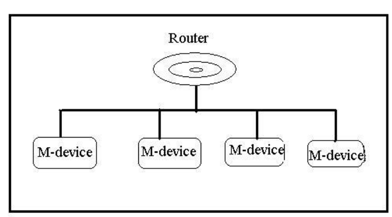


Fig. 1. Ad-hoc network (mobile nodes)

II. ROUTING PROCESS

Routing is a process of taking data over the network from source to a defined destination. Routing operate on layer 3 of the OSI model. Routing is almost defined with switching. The main difference between routing and switching is that routing operates at layer 3 and switching operates at layer 2 of the OSI model.

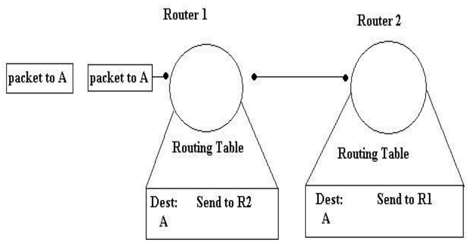


Fig. 2. Routing process from source to destination

Using both switching and routing mechanism with different information the whole process moves the data from source to destination. Routing process is different than switching process. In routing process when the source router sends the information to the neighbor router, then the neighbor router checks the route of the destination router in the routing table, if the route is available in the routing table, then it will send the information, if not then the router will discard the packet. If the router has more than one route available in the routing table then router select the best available path to the destination and sends the information [3].

RELATED WORK:

In this section the followings routing protocols are discussed.

A. Ad-hoc On-Demand Distance-Vector protocol (AODV)

AODV is a hop by hop routing protocol or in other words AODV is an on demand distance vector routing protocol. It has some features of DSDV protocol, for example using hop by hop, periodic notification messages and sequence numbers [4]. It can be seen as an updated DSDV protocol. It can also be seen as an updated DSR protocol. By means of an updated DSDV, it reduces the amount of broadcasting messages and only creates routes on need basis [5]. When a mobile node needs to send some information/data, then AODV find out a route to a destination, and it keeps the routes in the routing table up to the time, when they are needed by the source. In AODV the sequence numbers guarantee the loop free and freshness of routes in the routing table. AODV is relatively the same as the bellman-ford distance vector algorithm but it does work in mobile environment. [5-15]

B. The Dynamic Source Routing Protocol (DSR)

In [7] The Dynamic Source Routing Protocol (DSR) is a reactive unicast routing protocol. DSR is popular for some of its important features, which are, it is simple, dedicated to ad-hoc networks and very efficient. DSR has two methods for communication, which are,

- Route discovery
- Route maintenance

C. Destination sequenced distance vector Protocol

DSDV in [1], [8] is a proactive routing protocol in ad-hoc network, which uses bellman-ford algorithm. By using bellman-ford algorithm in ad-hoc network, it increment the sequence number of the new entry in the routing table for each device in the network. In order to operate correctly, DSDV end device has to send its full routing table to all neighbours periodically and vice versa to update its own routing table by getting the latest information from neighbour. All the end devices in the network have to update the routing table as soon as they get any update from neighbour. DSDV uses sequence number as its routing table attribute [16]. The sequence number shows the updated information. A route with higher sequence number is favourable than lower sequence number. Higher sequence number shows most updated information. If the two routes have same sequence number then the route with lower hop count will be preferred [2]. The sequence number is incremented with each broadcast. If there is any broken link the sequence number is tagged as infinity.Optimized Link State Routing Protocol (OLSR)

D. Optimized Link State Routing Protocol

Optimized link state Protocol (OLSR) is a proactive routing protocol. OLSR is the updated version of link state routing protocol. This means that the active routing paths will always be available in the routing table, if any mobile device needs them for communication. As soon as the topology gets changed, then every device sends a full routing table to all other mobile devices in the network. This will create an over head and bottleneck on the actual link. In order to reduce the overhead created by a big pile of broadcast messages, there is a technique used to reduce these broadcast messages. A network protocol uses Multipoint Relays (MPR). The basic job of MPR is to reduce the broadcast messages in some areas in the network and also to provide the shortest path [7]. According to [14] OLSR is an independent routing protocol, which does not have a fixed central administration and perform flat routing. OLSR is proactive routing protocol which requires all nodes have full updated routing information in the network. On the other hand the limitation of OLSR can be that it sends the updated information across the network which use a lot of the link bandwidth. But it has still minimized the flooding by the selection of MPR, which are only allowed to advertise Hello message. By changing the time interval between the broadcast timing the protocol can be more suitable for ad-hoc network. OLSR is very easy to be integrated in the existing operating system without changing header of IP.

III. SIMULATION RESULTS

Mobility Sequence:

The mobility file is generated using NS2 set-dest script. This model used by set-dest is changing position in mobility model. The model imposes a randomly motion, which a node move towards a different destination with a speed varying between zero and high speed parameter, while at the same time generating the mobility file. After stopping at this different destination for a specified 'pause time', the node continues this changing motion and stopping at a different destination until the simulation come to an end. The pause time parameter controls the motion of the node and is therefore a measure of mobility. For this reason, the pause time is varied to see its total effect. The selected pause times for this simulation are 10s, 15s, 20s, 30s, 50s, 100s and 110 sec. All parameters used to generate mobility file along with pause time is shown in the following table below,

No of nodes	30
Pause time	10, 15, 20, 30, 50, 100, 110
Maximum speed (m/s)	20
Simulation Time (s)	110
Area – X,Y	1500,300 (rectangular)

Table 1

A. *Traffic pattern*

Traffic files have been generated using cbrgen.tcl script which is part of NS2. Constant bit rate (CBR) traffic sources have been used. The parameters used for the traffic files are shown below in the table,

No of nodes	30
Seed	1
Maximum connections	10
Rate (Packet per second)	2.0

Table 2

Performance Metrics

The three performance metrics have been counted and plotted against the pause time. The results of the simulation are shown in the following graphs along with a detailed discussion.

Packet delivery ratio

Packet delivery ratio is low for both AODV and DSR at lower pause time i.e. when the motion is too high. Higher mobility causes often route breaks which means more route discoveries are made in case of reactive protocols. With lower mobility the route breaks are not very often which results in few route discoveries and hence better performance for reactive protocols. Between AODV and DSR, it is clear that AODV outperforms DSR in packet delivery ratio in case of high motion/mobility. The fact is that AODV uses fresh routes each time in case of route failure while DSR has route caching feature which means multiple routes to a destination are maintained. After one route fails, the other routes in tried instead of trying to discover another one. In case of high mobility, link breaking is often occur so chances for stale routes in DSR routing cache is high which is obvious from the results [17]. DSR route caching has a positive effect at lower mobility as shown in the graph since they are not very often route failures. Figure 4 shows the packet delivery ratio below,

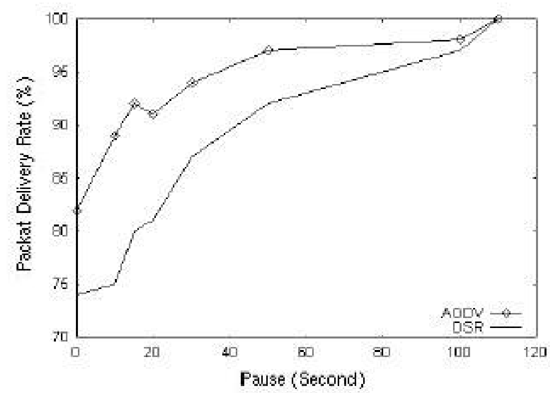


Fig. 4 Packet Delivery Rate vs. Pause Time

B. *Average Delay*

Average delay of AODV was higher than DSR at low pause time i.e. high mobility. This is because AODV generates more routing packet for discovering new routes in case of route failure which consumes bandwidth and therefore contributes to the delay in the network. On the other hand, DSR is utilizing route caching ability and making less route discoveries in case of route failures thus using little bandwidth and therefore delay is low for DSR. But the difference between the two is not much even though DSR is using route caching. The reason for this is that when stale routes in DSR cache are chosen it adds to the delay as well as to the bandwidth utilization and delivery time is wasted. As the pause time is increased i.e. mobility is decreased, the average delay for both AODV and DSR starts decreasing. Both start performing better with low mobility with AODV matching DSR at pause time of 110. Figure 5 shows the plotted graph below,

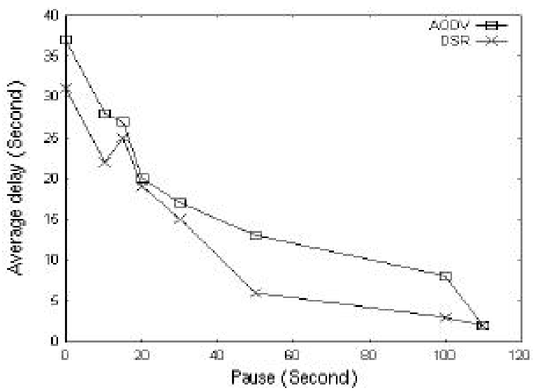


Fig. 5 show Average Delay vs. Pause Time

C. Average Routing Load

Average routing load for AODV is higher than DSR. The difference is high at lower pause time i.e. high mobility/motion. The reason for high overhead of AODV is the often route request packets for route discoveries which send this to every mobile node in the network. With high mobility this overhead is very high for AODV which relies upon fresh routes. DSR produces less overhead than AODV by utilizing route caching feature and using non-propagating route request packets for route discovery. With high pause time i.e. with lower mobility. The difference reduces with the decreased mobility. Figure 6 below shows the average routing load.

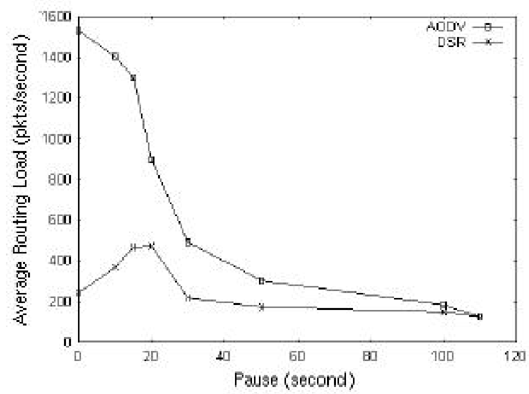


Figure 6 shows Average Routing Load vs. Pause Time

IV. CONCLUSIONS

Many routing protocols designed for mobile ad hoc networks are proactive, reactive or the combination of the two (hybrid). The proactive derives their mechanism from traditional fixed line network to the ad hoc networks. These protocols maintain a table of all the routes in the network. These protocols have high routing overhead in maintaining and updating these tables and are therefore recommended in situation where bandwidth is not a problem. While typical mobile ad hoc networks have bandwidth constraint, reactive routing protocols were designed for mobile ad hoc networks to address the issues in proactive protocols and conserve the bandwidth and power. Being reactive in nature, these protocols adapt to topological changes better than proactive protocols and are therefore best suited for mobile ad hoc networks.

PROPOSED SOLUTION

In a general perspective about the reactive routing protocols, it is clear from the results below that both the protocols perform very well under high pause time i.e. low mobility but their performance tends to degrade at higher mobility. This is due to the fact that high mobility often results in route failures which mean often route discoveries will be made by these protocols due to their reactive nature. Performance wise the results showed that

- AODV performed better than DSR in terms of packet delivery ratio

The performance gap was high at low pause time (high mobility) but with high pause time (low mobility).

- DSR started performing better and the gap was significantly reduced.

In terms of average end-to-end delay,

- DSR performed well with lower delay than AODV with at high mobility.
- DSR outperforms the AODV at high mobility with a high performance gap.

This is because AODV uses more route requests than DSR. The reason is that these route requests propagate to all the mobile nodes in the network. The low overhead of DSR is due to the route cache feature and non-propagating route requests.

Both protocols have their advantages and disadvantages in terms of different metrics and scenarios. The prime reason for low performance of AODV relies on a single route and at high mobility this results is often route requests. This can be overcome with a route caching technique to maintain multiple routes to a destination. However, on the other hand the route caching technique has negative impact on the performance of DSR at high mobility. At high mobility, the probability of stale routes in cache is high which degrades the performance. If some sort of cache route expiry mechanism is implemented than it would eliminate the probability of stale routes and thus would improve the performance of DSR and AODV can also benefit from a similar caching technique.

V. THE FUTURE

Today's ad-hoc research and studies have proved that in future the mobile router will provide the internet connectivity to ad-hoc end nodes. Networks will be allowed to move freely to different locations, but there will not be any impact on the actual link/connectivity. Mobile ad-hoc networks are of more interest for different type of applications, such as distributed sensing networks, distributed collaborative computing. Future need for ad-hoc networking is the field of mobile computing, with its current focus on mobile IP function, will expend significantly. In the future, ad-hoc will require a high level communication technology to get multi-hop clusters that will be able to function independently. Research in GPS system will also improve the performance of mobile ad-hoc networks.

REFERENCES

[1] Working Group IEEE 802.11, April 2008. [Online] Available: http://www.ieee802.org/11.

[2] Freebersyser and B. Leiner, ed. C. E. Perkins, Addison-Wesley "A DoD Perspective on Mobile Ad-hoc Networks," Ad-hoc Networking, pp.29-51, 2001.

[3] R. Perlman, "Network Layer Protocols with Byzantine Robustness," Ph.D. Dissertation, MIT/LCS/TR-429, MIT, pp.77-82, 1988.

[4] PERKINS, C. E. BHAGWAT, "Highly dynamic destination-sequenced distance-vector routing for mobile computers". ACM SIGCOMM, Proc. of the conference on Communications architectures, protocols and applications, 2000.

[5] Agustin Zaballos, Alex Vallejo, Guiomar Corral, AdHoc routing performance study using OPNET Modeler Agustin Zaballos, Alex Vallejo, Guiomar Corral, Jaume Abella University Ramon Llull (URL-La Salle Engineering), Barcelona (Spain), 2004.

[6] Das S. Perkins C.E., Belding-Royer E.M. Ad-hoc on-demand distance vector (aodv) routing. RFC 3561, IETF Network Working Group University of Cincinnati July 2003.

[7] David B. Johnson David A. Maltz Josh, Broch "The Dynamic Source Routing Protocol for Multi-Hop Wireless Ad-hoc Networks" Computer Science Department Carnegie Mellon University Pittsburgh, PA 15213-3891, 2007.

[8] S. Murphy, "Routing Protocol Threat Analysis," Internet Draft, draft-murphy-threat-00.txt, October 2002.

[9] T. Clausen and P. Jacquet "Optimized Link State Routing Protocol (OLSR)." RFC 3626, IETF Network Working Group, 2008.

[10] Scalable Routing Protocols for Mobile Ad-hoc Networks Xiaoyan Hong, Kaixin Xu, Mario Gerla Computer Science Department, University of California, Los Angeles, CA 90095, 2008.

[11] Aleksandra Huhtonen "Comparing AODV and OLSR Routing Protocols", Helsinki University of Technology Telecommunication Software and Multimedia Laboratory, 2007.

[12] Tony Larsson and Nicklas, Hidman Routing protocols in wireless ad-hoc networks, 2007.

[13] Perkins, C., Royer, E. M. and Das S., 2000. Performance comparison of Two On-Demands routing Protocols for Ad-Hoc Networks, IEEE INFOCOM 2000.

[14] ProtocolKalyan Kalepu, Shiv Mehra and Chansu Yu, "Experiment and Evaluation of a Mobile Ad Hoc Network with AODV Routing" Department of Electrical and Computer Engineering Cleveland State University 2121 Euclid Avenue, SH 332, Cleveland, OH 44115, 2004.

[15] Yao, Yanjun, Qing Cao, and Athanasios V. Vasilakos. "EDAL: An energy-efficient, delay-aware, and lifetime-balancing data collection protocol for heterogeneous wireless sensor networks." *Networking, IEEE/ACM Transactions on* 23.3 (2015): 810-823.

[16] Hammoudeh, Mohammad, and Robert Newman. "Adaptive routing in wireless sensor networks: QoS optimisation for enhanced application performance."*Information Fusion* 22 (2015): 3-15.

[17] Kiani, Farzad, et al. "Efficient intelligent energy routing protocol in wireless sensor networks." *International Journal of Distributed Sensor Networks* 2015 (2015): 15.

Artificial Neural Network Approach for Stock Price and Trend Prediction

Nasimul Hasan, Risul Islam Rasel

Department of Computer Science and Engineering, International Islamic University Chittagong
Chittagong, Bangladesh
nasimul_hsn@yahoo.com, rasel.kmutnb@gmail.com

***Abstract*—Nowadays, Demand of forecasting stock market price is increasing at a higher rate than the ever before as more people are getting connected to the stock business. As many criteria play more or less strong inductive role over the stock market, the trend and price always keep changing here. So, it is challenging to predict exact price value. But some Data mining and Machine learning techniques can be implemented to do this challenging task to predict stock market price and trend. In this study, Artificial Neural Network (ANN) is used along with windowing operator; which is highly efficient for working with time series data for predicting stock market price and trend. This study is done on Wal-Mart Stores Inc. (WMT) a listed company of New York Stock Exchange. Five years historical dataset (2010-2015) is used to undertake the experiments of this study. According to the result of this study Artificial Neural Network (ANN) can produce a rational result with a small error.**

I. Introduction

The Stock market is a vital part of the economy of a nation. Money market plays a crucial role in the development of the business and trade of a nation that in the end influences the economy of the nation. This is the reason that the government, business organizations and even the national banks of a nation keep a close observation on the happenings of the share trading system. The share trading system is essential from both the business perspective and the financial analysis perspective. Thousands of people are getting involved in this potential business these days. It is a fundamental need for the people who are directly connected to the market to have an insight regarding the market trend [1] [2] [3]. So, forecasting stock price and market trend are getting more and more importance among the people. Stock market is basically a nonlinear, non-parametric, boisterous and deterministically disordered market [2] [3] [4]. The price and trend are frequently influenced by some critical and crucial factors; like liquid money, supply and the demand of goods, earnings of a company, the political situation etc. An Artificial Neural Network (ANN), generally called Neural Network (NN), is a scientific model or computational model that is propelled by the structure and features of natural neural systems. Artificial Neural Network (ANN) is widely used mostly in classification, regression, clustering, anomaly detection etc. A neural framework contains an interconnected assembling of artificial neurons, and it forms data utilizing a connectionist way to deal with reckoning (the focal connectionist rule is that mental phenomena can be portrayed by interconnected systems of straightforward and regularly uniform units). As a rule; ANN is a versatile framework that progressions its structure in light of outside or inner data that courses through the system in the learning stage. Current neural systems are normally used to model complex connections between inputs and yields or to discover patterns in information [9]. YETIS, KAPLAN, and JAMSHIDI [5] showed in their research that Artificial Neural Network (ANN) can give an appreciative result with a very low error rate. Their model produced a result of 99 percent accuracy, where the best validation performance (MSE) was 37.12 which mean the error found in that model was less than 2%. Phua, P. K. H. Ming, W. Lin [8] combined ANN with Genetic Algorithm and predicted the stock price with 81% accuracy. In this study an approach of combining Artificial Neural Network with windowing operator which is very efficient for time series data prediction has been proposed. Three effective and highly efficient models, Model 1 for 1 day ahead prediction, Model 2 for 5 days ahead prediction and model 3 for 10 days ahead prediction is proposed here in this study. Comparison between some other novel algorithms like Support Vector Machine (SVM) and K-Nearest Neighbor (KNN) is shown in Table VII to understand the performance difference.

II. Methodology

A. *Artificial Neural Network (ANN)*

The goal of this study is to enhance the precision of day by day stock value forecasting of securities exchange by utilizing the neural network. An ANN has a few points of interest yet a standout amongst the most perceived of these is the way that it can really gain from watching information sets. Along these lines, ANN is utilized as an arbitrary capacity estimate instrument [11]. These sorts of devices gauge the most effective and perfect systems for touching base at arrangements while characterizing figuring capacities or disseminations. ANN takes information tests instead of whole information sets to touch base at arrangements, which spares both time and money. ANNs are considered genuinely straightforward numerical models to improve existing information investigation innovations. ANNs have three layers that are interconnected. The primary layer comprises of data neurons. Those neurons send information on to the second layer, which thus sends the yield neurons to the third layer [12]. The study utilized three-layer (a hidden layer) perception model (a feed forward neural network) prepared with back propagation calculation. Authentic stock costs of distinctive organizations were taken from distributed stock information on the Web. The learning capacity or the initiation work that was utilized is sigmoid equation

$$f(x) = \frac{1}{(1 - e^{-fx})} \tag{1}$$

Neural Network gets various inputs (either from the unique information or from the yield of different neurons in the neural network). Every information comes through an association that has a quality (or weight); these weights relate to synaptic adequacy in an organic neuron. Every neuron additionally has solitary limit esteem. The weighted aggregate of the inputs is shaped, and the limit subtracted, to make the initiation out of the neuron (otherwise called the post-synaptic potential, or PSP, of the neuron). The activation signal is then passed through an activation function [6].

$$yk(x, \omega) = \sigma(\sum_{j=0}^{M} \omega_{kj}^{(2)} h(\sum_{i=0}^{D} \omega_{ji}^{(1)} xi)) \tag{2}$$

Here, yk is a set of output variables controlled by ω, which is adjustable parameter. The parameter $\omega_{ji}^{(1)}$ is the weights and $\omega_{kj}^{(2)}$ is the biases. The superscript indicates the position of layer [6].

B. Time Series Data

Time series data are a kind of data where the values of an attribute or variable are stored in such a way that the time interval for a value is exactly same in comparison with the previous and the next value. Time series data indicates the change of value over time. A time series can also demonstrate the effect of cyclical, seasonal and irregular events on the data item being measured.

C. Windowing Operator

This is a mechanism which changes a given sample set containing series data into another sample set containing single valued cases. For this reason, windows with a predefined window and step size are moved over the series and the characteristic quality lying horizon values after the window end is utilized as a label which ought to be forecast. This administrator can handle multivariate series data too.

D. Evaluation Processes MAPE:

Mean average percentage error (MAPE) is a measure of exactness of a system for developing fitted time arrangement values in insights, particularly in pattern estimation. It ordinarily communicates precision as a rate and is characterized by the technique [7].

$$MAPE = 100 \frac{\sum_{i=1}^{n} \left| \frac{A-P}{A} \right|}{n} \tag{3}$$

Here, A defines actual price, P defines the predicted price and n defines the number of days calculated.

E. Evaluation Processes RMSE:

Root mean square error (RMSE) is a famous evaluation process to calculate the error rate of a regression model. Though, it can only be compared between models with errors calculated in same units.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (yt - \hat{y}t)^2}{n}} \tag{4}$$

Here, yt is the original value of a point for a given time period t, n is the total number of fitted points, and n is the fitted forecast value for the time period t.

III. Experiment Design

A. Research Data

The proposed model can produce a rational result for almost every company. For a convenient study, a well-known company is considered here. The 5-year historical data (2010-2015) of Wal-Mart stores Inc., a listed company of New York Stock Exchange was considered for the experiment and evaluation in this study. The number of instances of the data set is 10805. Table 1 shows the attribute merit and rank measured by I_{GAIN} using 5 fold cross validation. In this study, the main aim was to predict the closing price of a Stock. For that purpose, three models are proposed here. They are 1 day ahead, 5 days ahead and 10 days ahead model. Five attributes; Date, Open price, Close price, High price and Low price are used in this study. The attribute 'Date' was chosen as id and the attribute 'Close price' as the label. The rest were kept as regular attributes. The dataset was divided into two parts. 80% of the data (2010-2014) were taken as training data and the rest 20% (2014-2015) as test data. Figure 1 shows the sample dataset for the year of 2015. Here the X axis denotes the price in BDT and the Y axis denotes the corresponding dates.

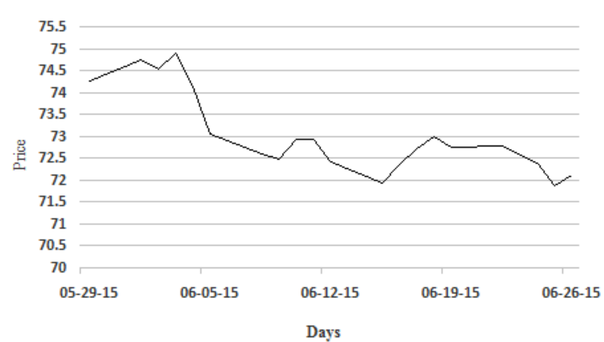


Fig. 1. Actual Price of WALMART INC. stock, 2015

B. Data Preprocessing

Five years historical data of Wal-Mart Inc. from New York Stock Exchange was considered for this study. The attributes needed for the study were chosen carefully. Then we run a very important preprocessing called Missing value handling to find out and replace the missing values. In this study an

TABLE I
ATTRIBUTE SELECTION 5 FOLD CROSS-VALIDATION, SEED: 1

Attribute	Average Merit	Average Rank
High	2.526 ± 0.003	1 ± 0
Open	2.503 ± 0.004	2 ± 0
Low	2.455 ± 0.003	3 ± 0
Volume	0.336 ± 0.006	4 ± 0

attribute Date is used as id and Neural Net can only handle numeric data. So, an operator Adjust date was used in order to convert the date type value into numbers.

C. Windowing Operator Analysis

The next process was the use of windowing operator to change time series data into generic data. Table II shows the windowing operator analysis for the result produced in this study. The parameter ’Windowing Size’ denotes the size (Number of example for training) of the training window. The ’Step size’ is the number step the window moves forward.

TABLE II
WINDOWING OPERATOR ANALYSIS

Model	**Windowing Size**	**Step Size**	**Training Window Width**	**Training Step Size**	**Testing Window Width**
1 day ahead	3	1	2	1	2
5 day ahead	3	1	3	1	3
10 day ahead	3	1	3	1	3

D. Neural Net Function Analysis

The learning stage begins with the application of Artificial Neural Network (ANN). For the best result, the function and the parameters of ANN were chosen carefully. Here, α =learning rate and M=Momentum. Table III shows the function settings of Neural Net used in this study.

TABLE III
NEURAL NET FUNCTION SETTINGS

Model	**Training cycle**	α	**M**	**Error Epsilon**
1 day ahead	1300	0.3	0.2	1.0E-5
5 day ahead	1300	0.3	0.2	1.0E-5
10 day ahead	1300	0.3	0.2	1.0E-5

E. Sliding Window Validation

For this study a special validation process, Sliding window validation was applied. This is a unique approval chain which must be utilized for series forecasting where the time focuses are encoded as cases. It utilizes a certain window of cases for preparing and uses another window for testing. The window is moved over the case set and the average is determined of all execution estimations. The parameter ’cumulative training’ shows if every single previous sample ought to be utilized for preparing (rather than just the present window) [10]. Table IV shows the property settings of the validation process (Sliding Window validation) used in this study.

TABLE IV
VALIDATION PROCESS PROPERTIES

Properties	**1 day ahead**	**5 day ahead**	**10 day ahead**
Training Window Width	2	2	1
Training Window Step	1	1	1
Test Window Width	2	2	1
Horizon	1	5	10
Cumulative Training	No	No	No

F. Model Setting and Analysis steps

The experimentation models are begun with data preprocessing steps to deliver inputs to ANN. For that, windowing technique, for example, rectangular windowing was utilized as data preprocessing strategies. Windowing operator changes the time series information into a universal dataset into the learning process [1] [2] [3]. In this study, the Artificial Neural Network (ANN) was utilized as a learning algorithm to understand the trend pattern from the dataset and to anticipate the stock cost as yield in view of that learning. This study is led in two stages, training stage, and testing stage. Steps from these two stages are given below:
1) Training stage
Step 1: Read the training data.
Step 2: Adjust Date.
Step 3: Apply windowing operator to transform the time series data into a generic dataset. This step will convert the last row of a windowing within the time series into a label or target variable. Last variable is treated as label.
Step 4: Perform a sliding windowing validation process of the produced label from windowing operator in order to feed them as inputs into ANN model.
Step 5: Select training cycles and special parameters of ANN (learning rate, momentum, error epsilon).
Step 6: Run the model and observe the performance (accuracy).
Step 7: If the accuracy is good than go to step 8 or go to step 4.(As the main motive of the study was to improve accuracy of Stock Price prediction, the best parameter combination should be set. So, if the result found in this step is not good enough; the whole process should be done from step 4 again)
Step 8: Exit from the training stage and apply trained model to the testing dataset.
2) Testing stage
Step 1: Read the testing dataset.
Step 2: Apply the training model to test data
Step 3: Produce the predicted price and market trends

Figure 2 shows the experiment process in flowchart.

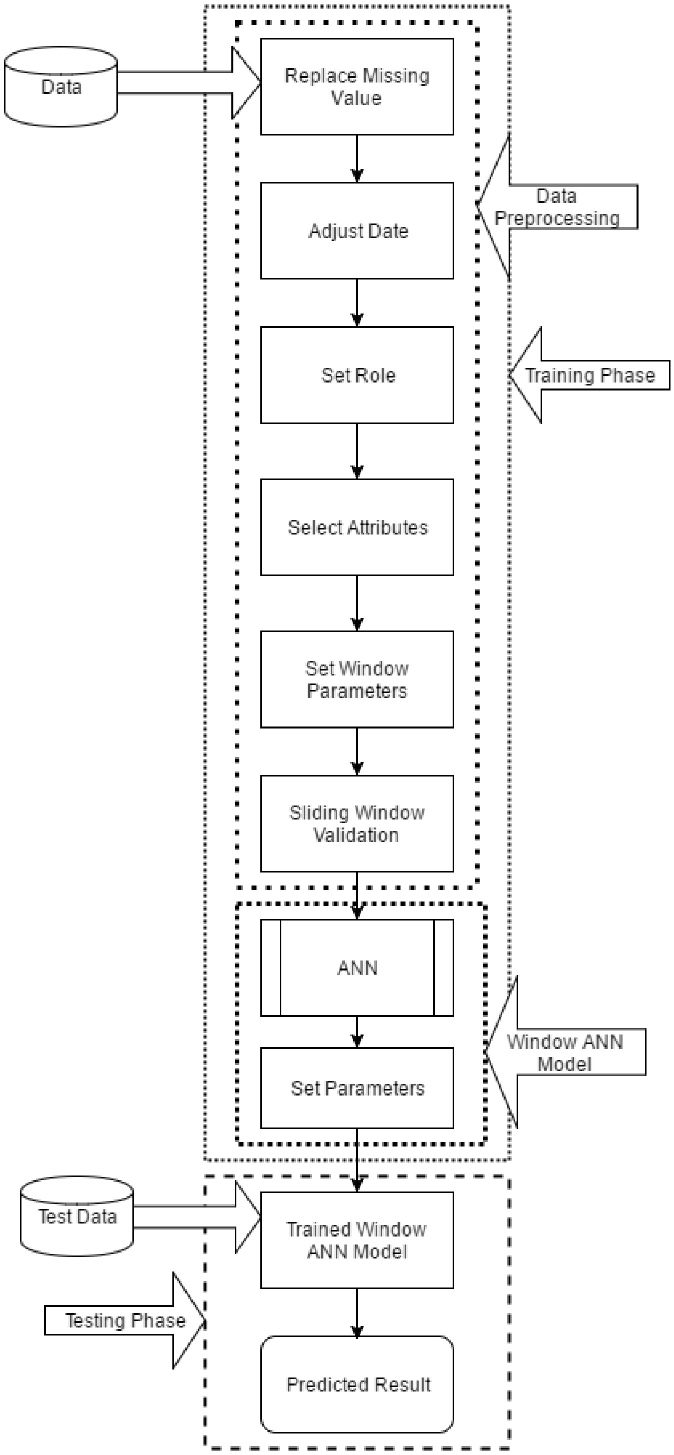


Fig. 2. Experiment Setting Flowchart

IV. Experiment Result

Table VI displays the predicted stock price by the proposed ANN models. All the three model predicted almost perfectly. Table VII shows the comparison of different algorithms on the same dataset for all three models.

TABLE V
NEURAL NETWORK MODEL

Hidden Layer 1	
Threshold: 0.185	
Node 1	0.665
Node 2	-0.561
Node 3	-0.539
Node 4	-1.005
Node 5	-0.532
Node 6	0.619
Node 7	0.652
Node 8	0.679

TABLE VI
RESULT FOR PROPOSED ANN MODEL

Date	Actual Price (USD)	Predicted Price		
		1 day ahead	**5 day ahead**	**10 day ahead**
13-01-15	89.30	89.44	87.90	91.71
07-11-14	78.76	77.63	77.72	78.08
29-07-14	75.44	75.00	76.43	77.59
04-06-14	77.12	76.26	76.96	77.53

A. Error Calculation (MAPE & RMSE)

The error is calculated between the actual price and the predicted price generated by the ANN model. Two evaluation processes, Mean Average Percentage Error (MAPE) and Root Mean Square Error (RMSE) are used in this study to find out the error of the models. Table V shows the MAPE and RMSE for the models which were applied only on the testing data.

B. Graphical Representation of the Study

In this study, three different models are proposed for forecasting stock market price and trend. Different values for all the parameters were used to get the best result for each model.Figure 3 shows the correlation between the attributes (or feature). The correlation is determined by using normal correlation function. The scale 1 to -1 denotes the level of relation. 1 means strong relation and -1 denotes very weak relation. Figure 4, Figure 5 and Figure 6 shows the graphical representation of the difference between actual price and predicted price for 1 day, 5 day and 10 day prediction respectively.

V. Conclusion

A. Discussion

The motive of the study was to construct an effective and an efficient model to forecast stock price and stock market trend using Artificial Neural Network along with some special operators with meaningful selection of parameters of the operators. Two different evaluation processes, MAPE and RMSE were used to calculate the rate of error and the proposed models are capable of predict stock market price and trend with very little error. 1 day ahead model which predicts the price of 1 day ahead market can predict the best among the

TABLE VII
COMPARISON OF DIFFERENT ALGORITHMS

Dataset	Model	Horizon	ANN		SVM		KNN	
			MAPE	RMSE	MAPE	RMSE	MAPE	RMSE
Walmart Inc.	1 day ahead	1	**0.75**	**0.60**	2.57	1.90	2.71	2.28
	5 day ahead	5	3.28	2.73	0.41	0.33	3.40	2.82
	10 day ahead	10	**2.01**	1.59	2.07	1.56	4.47	3.70

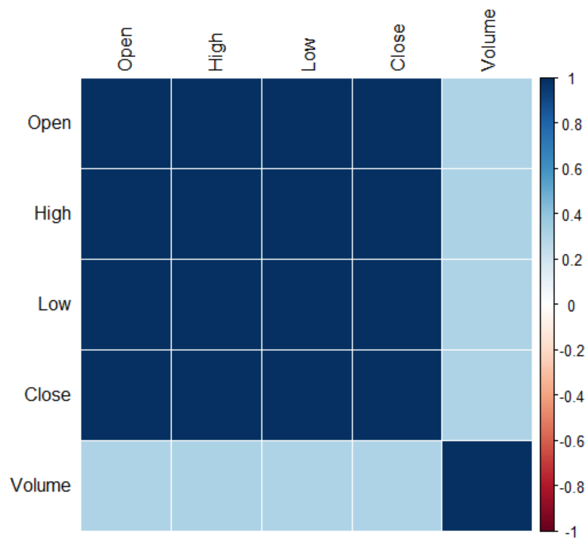


Fig. 3. Correlation between the features

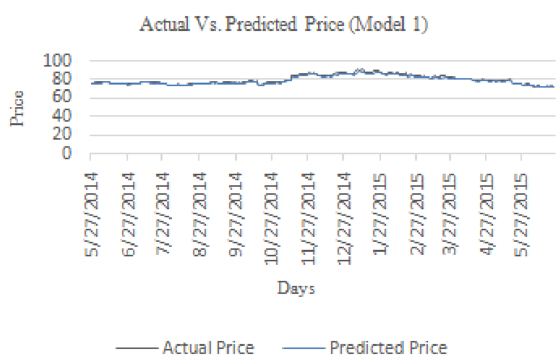


Fig. 4. Actual Price vs. Predicted Price, 1 day Model

three models. The proposed model can be used to predict stock price and trends instead of current techniques with low accuracy and thus it can help the business related people as well.

B. Limitation and Future Work

Only Windowing operator was used in this study for data preprocessing step and the study was designed based on only the New York Stock Exchange. In future other data preprocessing techniques will be used. Some other algorithms will be applied to determine the best model and to upgrade

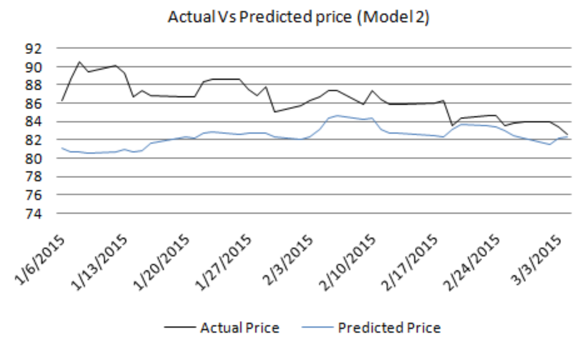


Fig. 5. Actual Price vs. Predicted Price, 5 day Model

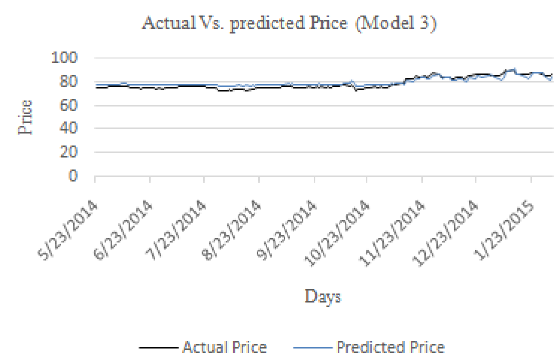


Fig. 6. Actual Price vs. Predicted Price, 10 day Model

prediction result. Different dataset from different stock markets will be applied in order to form a universal model for every market.

REFERENCES

[1] Meesad, Phayung, and Risul Islam Rasel. "Dhaka Stock Exchange Trend Analysis Using Support Vector Regression." The 9th International Conference on Computing and InformationTechnology (IC2IT2013): 9th-10th May 2013 King Mongkut's University of Technology North Bangkok. Vol. 209. Springer Science & Business Media, 2013.

[2] Meesad, Phayung, and Risul Islam Rasel. "Predicting stock market price using support vector regression." Informatics, Electronics & Vision (ICIEV), 2013 International Conference on. IEEE, 2013.

[3] Rasel, Risul Islam, Nasrin Sultana, and Phayung Meesad. "An efficient modelling approach for forecasting financial time series data using support vector regression and windowing operators." International Journal of Computational Intelligence Studies 4.2 (2015): 134-150.

[4] Lu, Chi-Jie, et al. "Stock index prediction: A comparison of MARS, BPN and SVR in an emerging market." Industrial Engineering and Engineering

Management, 2009. IEEM 2009. IEEE International Conference on. IEEE, 2009.

[5] Yetis, Y.; Kaplan, H.; Jamshidi, M., "Stock market prediction by using artificial neural network," in World Automation Congress (WAC), 2014 , vol., no., pp.718-722, 3-7 Aug. 2014

[6] Ince, H., Trafalis, T.B.: Kernel Principal Component Analysis and Support Vector Machines for Stock Price Prediction, pp. 20532058 (2004)

[7] Lucas, K., Lai, C., James, N., Liu, K.: Stock Forecasting Using Support Vector Machine. In: Proceedings of the Ninth International Conference on Machine Learning and Cybernetics, pp. 16071614 (2010).

[8] Phua, P. K. H. Ming, W. Lin, "Neural Network With Genetic Algorithms For Stocks Prediction", Fifth Conferences of the Association of Asian-Pacific Operations Research Societies, Singapore, 5th-7th July, 2000

[9] Yunus YETIS , Halid KAPLAN, Mo JAMSHIDI, Stock Market Prediction by Using Artificial Neural Network, In World Automation Congress, 2014

[10] F.R.EngLucas, K., Lai, C., James, N., Liu, K.: Stock Forecasting Using Support Vector Machine. In: Proceedings of the Ninth International Conference on Machine Learning and Cybernetics,pp. 16071614 (2010)

[11] Chan, C. C., E. W. C. Lo, and Shen Weixiang. "The available capacity computation model based on artificial neural network for leadacid batteries in electric vehicles." Journal of Power Sources 87.1 (2000): 201-204.

[12] Kim, Kyoung-jae, and Ingoo Han. "Genetic algorithms approach to feature discretization in artificial neural networks for the prediction of stock price index." Expert systems with Applications 19.2 (2000): 125-132.

Sensorless Temperature Monitoring System using GSM Module for Smart Home Applications

[1]Md. Mahamudul Hasan, [2]Md. Yeasin Arafat, [3]Mahtab Murshed and [4]M. Abdur Razzak

Department of Electrical & Electronic Engineering, Independent University, Bangladesh
Plot-16, Block-B, Bashundhara, Dhaka-1229, Bangladesh
E-mail: [1]hemal.medio366@gmail.com, [2]nealmagik@gmail.com, [3]mahtab_murshed@hotmail.com, [4]razzak@iub.edu.bd

***Abstract*—Smart Home innovation began for over 10 years to present the idea of system administration gadgets and gear in the house. As indicated by the Smart Homes Association, the best meaning of keen home innovation is: the mix of innovation and smart technologies through home networking for a superior nature of living. Numerous instruments that are utilized as a part of PC frameworks can likewise be incorporated in Smart Home Systems. Global standard for mobile communications (GSM) is the most operated networking system till to date with its cheaper accessibilities and keen communication possibilities. The development, testing, and use of sensorless temperature monitoring for smart home embedded system that utilizes GSM networking module are presented in this paper for remotely home temperature information accessing and setting intended temperature threshold. The system also autonomously makes sure of working reliability using GSM USSD gateway.**

Keywords—GSM modem, temperature monitoring, smart home, embedded system

I. Introduction

Nowadays, monitoring of different environmental parameters like temperature, humidity, light intensity etc. are more admired in different field. Even to get more accuracy and flexibility on that monitoring technique much sophisticated systems were invented. Temperature is a target comparative measurement of hot and cold. A sudden change of temperature can cause a great amount of impact on the home and its dwellers. Excessive hot temperatures can cause various property damages. Besides, improper heat can cause illness when the temperature reaches above 90 degrees Fahrenheit. Inhabitants, especially pets, can be vulnerable to heat related sickness as well. These issues can be overcome significantly by developing and proper usage of convenient smart embedded systems. There are some other concerns can be realized like when dwellers get away from the house in a busy manner the managing of heating and air conditioning system can be useful to improvise the proper usage of energy. With remote temperature control system, an optimized energy use can be made sure by cutting back on furnace or air conditioner use in an empty house [1-4].

A system comes with temperature tolerance level should not be exceeded that by any means to produce expected outputs. But, in real world that can be caused intentionally or unintentionally. In smart home application, there are many systems which are composed of temperature sensor, ranging from -55 degree Celsius to 138 degree Celsius temperature to monitor and report necessarily though GSM module or other communication device from the system. On the contrary, the other parts of these systems especially communication devices, which play a vital role in reporting, cannot have that much of tolerance level in comparison with that temperature sensor's sensing ability. So, a system has been realized which can utilize the communication module as a standard home temperature monitoring unit within its temperature tolerance range and serve as a smart system as well allowing remotely temperature access and setting up.

Different sort of sensors, microprocessors, complicated network topology and different types communicational module are used for sensing, processing and transmitting the value of these parameters in real time [1-4]. Temperature sensing and its data processing are mostly usable environmental parameter in most of the systems. Basically the behaviors of electronics devices fully depend on changing of temperature in all types of machine. Even rapid changing in temperature can damage the ideal working behavior of any devices. So measurement of temperature and check its unusual behavior over real time is a big concern. Also real time temperature data checking over wireless communication system remotely is another vital concern.

To minimize these concerns, different types of system were proposed or designed earlier using different types of temperature sensors. Few research works shown CMOS based RFID temperature sensor, LM35 or AD590 temperature sensor based measurement system [2, 5-7]. Although the CMOS based temperature sensor is only used in passive RFID tag and suitable for human body temperature measurement. On the other hand, LM35 or AD590 temperature sensors are basically used for room, environment or soil temperature measurement system. Whatever the sensor are used for measuring temperature on any system, GSM module is used for wireless communication of remote data accessing in every system [1, 4, 5, and 7].

In smart home systems, the use of temperature measurement and flexible data accessing through GSM module are basic criteria. That's why in every smart home system temperature sensor, data processor, and GSM module are basic components. Though, system reliability is one of the major problems for these types of sophisticated systems. However, the proper use of GSM module can precisely remove the external temperature sensor from the smart home system applications.

To reduce these problems and increase the system reliability, a sensorless temperature monitoring system for smart home embedded application using GSM module is proposed and implement in this research work.

II. PRELIMINARIES

The block diagram of the proposed system is shown in Fig.1. To make a clear understanding about this research work, some important issues which will be used in the development of this system are discussed in this section.

A. *SIM Card*

Subscriber Identity Module (SIM) card can be seen as very tiny computer as it has central processing unit, electronically erasable, programmable read only memory (EEPROM) which contains the user data, random access memory (RAM) for program execution, and read only memory (ROM) for the operating system, user authentication and data encryption algorithms, and other applications [8]. Besides, memory parts of SIM have a trend in increasing storage capacity. Accessing this memory by AT commands, SIM card storage can be dynamically controlled. It is also noted that 30-40 SMS can be stored in SIM chip presently available in Bangladesh [9].

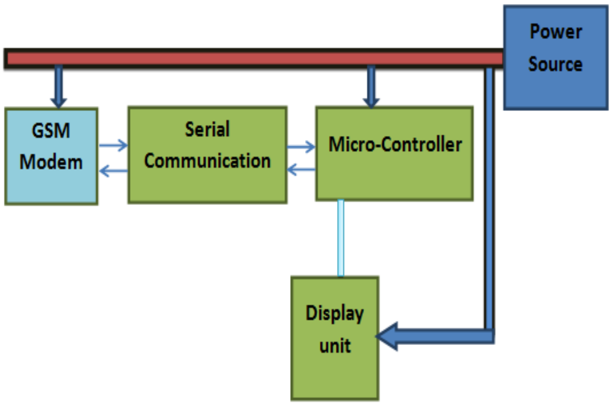


Fig. 1. Total system block diagram

TABLE I. IMPORTANT AT COMMANDS AND ITS UNDERSTANDING

AT Commands	**Understanding**
AT+CUSD	Writing USSD code
AT+CMGF=1	Set message format to TEXT mode
AT+CSAS	Save SMS settings. This may take up to 10 seconds
AT+CPMS	Preferred storage
AT+CMGR=1	Read SMS message in SIM location 1
AT+CMSS=2	Send message from memory location 2
AT+CMGD=1	Delete SMS message in SIM location 1

B. *AT Command*

AT commands have been developed by Dennis Hayes for modem interaction language and by time, the ETSI GSM 07.07 (3GPP TS 27.007) has directed its command style for controlling a GSM phone or modem [10]. This language is used to talk to GSM modem. The AT commands of ESTI GSM 07.07, GSM 07.05 and SIM technology company-SIMCom extended standard, those will be used in the proposed system, are listed in Table I..

C. *USSD Gateway*

USSD is a protocol used by GSM cellular telephones to communicate with the service provider's computers and its gateway routes USSD messages from the signaling network to a service application and back [11]. To access network operator services, such as, balance refill or inquiry, special offer, internet package and so on, USSD codes can be used. These codes differ with local mobile network operator service providers. The USSD codes listed in Table II, which have been adapted in this proposed work, are the codes of "Banglalink" - a local mobile network operator service provider in Bangladesh.

TABLE II. USED USSD CODE AND ITS CORRESPONDING AT COMMAND

USSD Code	*124#	Banglalink balance check
AT command	AT+CUSD=1, '*124#'	Accessing service

III. TEMPERATURE SENSING APPROACH

The proposed approach incorporates SimCom GSM/GPRS communication chip such as SIM300, SIM900, SIM900A, SIM908 and so on for temperature detection since, nowadays, it is being widely used in GSM modem based embedded system application design. This section will illustrate the temperature data extracting, processing, and testing. In order to perform these operations, any module with SimCom communication chip needs to go through a setup process which can be done by AT command. The block diagram for determining the temperature through the AT commands is shown in Fig. 2.

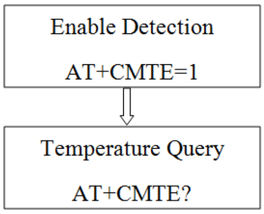


Fig. 2. AT command writing to ME

After writing temperature query command to Mobile Equipment (ME), where communication chip is attached, the ME sends response with the style, +CMTE: 1,Temp and ended with OK string. The Temp (temperature) parameter varies from -30 to 80 degree Celsius though the chip can capture -40 to 90 degree Celsius [12]. Now, the ME response data can be processed in the application processor for only temperature data extraction. The logical flowchart for temperature data extraction is given in Fig. 3.

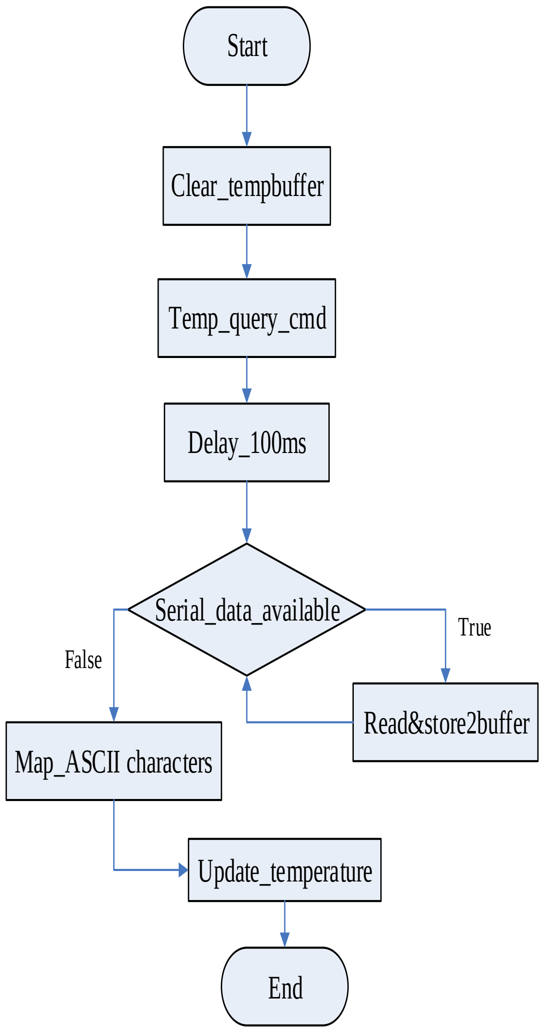


Fig. 3. Flowchart for temperature extraction

IV. IMPORTANT SYSTEM'S DEVICES

The proposed work offers a simple system with less number of components to measure the temperature remotely. In this section the important devices used in the proposed system were discussed.

A. GSM Module

In the development process, a SIMCOM SIM900 GSM shield has been integrated in the system. SIM900 uses AMR926EJ-S chip processor which is very powerful. This module can be controlled by different version of AT commands set such as- GSM 07.07, GSM 07.05 and SIMCOM enhanced AT commands. This module uses 1.5mA current in sleep mode and 3.1v to 4.8v as supply voltage. The operating temperature is -40 degree Celsius to 85 degree Celsius.

B. Processor

In this work, Arduino ATmega328 microcontroller has been adopted for the purpose of whole system processing. It is an Atmel 8-bit AVR RISC-based microcontroller which combines 32 KB ISP flash memory with read-while-write capabilities, 1 KB EEPROM, 2 KB SRAM, 23 general purpose I/O lines, 32 general purpose working registers, three flexible timer/counters with compare modes, internal and external interrupts, serial programmable USART, a byte-oriented 2-wire serial interface, SPI serial port, 6-channel 10-bit A/D converter (8-channels in TQFP and QFN/MLF packages), programmable watchdog timer with internal oscillator, and five software selectable power saving modes. The device operates between 1.8-5.5 volts. The device achieves throughputs approaching 1 MIPS.

C. Monitor

Nokia 5110 - a basic graphic LCD having the same parameters and applications of the Nokia 3310 LCD, which uses the PCD8544 controller, is used in the proposed system. The PCD8544 is a low power CMOS LCD controller/driver and is designed to drive a graphic display of 48 rows and 84 columns. All necessary functions for the display are provided in a single chip, including on-chip generation of LCD supply and bias voltages, resulting in a minimum of external components and low power consumption. The PCD8544 interfaces to microcontrollers through a serial bus interface

V. SYSTEM OPERATION

The proposed system is designed as user interactive system. It is intended to take user input first through SMS to enable the interactive behavior. After setting the user data up, the system keeps watching to receive SMS from system's registered number for performing action accordingly. The system process the incoming SMS followed by table III which undergoes a specific SMS forwarding pattern.

The system extracts the valid number SMS information and network operator service information by mapping the ME responded ASCII positioning of incoming SMS.

TABLE III. SMS PROCESSING

Forwarding SMS style	System Understanding
1,0880-xxxxxxxxxx,50	1=SMS for set up, mobile number, alarming temperature
2,action temperature, turn on/off	2= SMS for temperature setting where the system will turn on/ off the load 1= turn on, 0=turn off
3	3=SMS for sending present temperature update

The system autonomously checks the system operational condition such as whether it will work or not. If the system gets out of network, balance shortage, and communication linkage then it will be updating by turning on a LED light for indication. But, if the system is such as, only getting out of data service charge then it will be updating the registered user by sending a SMS. The pseudo code of reliability is shown in Fig.4. The full system working flowchart is shown in Fig. 5.

```
void test_reliabilty {

Serial.write(communication command);
delay_ms(10);
R=serial.read();
        If R is equal to Ok
                Serial.write(signal strength command);
                S=serial.read();
                strength=find_strength(S);
                If s is good
                        Serial.write(USSD code);
                        delay_ms(1000);
                        B=Serial.read();
                        balance=find_balance(B);
                        If balnce is greater than 2 and less than 5
                                send_sms(registerd number);
        else
                Turn on LED light;
        End If
}
```

Fig. 4. Pseudo code of reliability

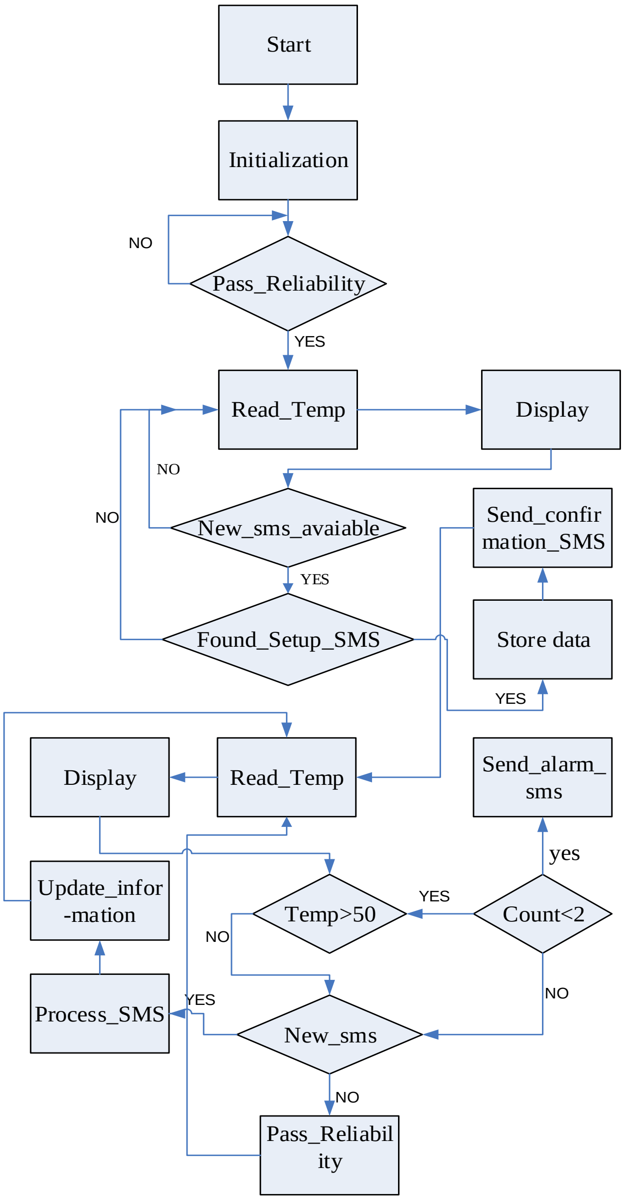


Fig. 5. Full system flowchart

Finally, the end user display unit will be displaying four types of data: current temperature, action temperature, warning temperature and system's status.

The current temperature data are collected from GSM modem and after registration action temperature that is, the temperature where system will turn on or off any electrical load, can be set by SMS. Warning temperature set by SMS plays an important role to warn the system's user by SMS when the system senses above warning temperature due to any abnormal home environment.

VI. Testing and Prototyping

To check the validation of the system simulation based testing and finally prototyping are examined.

A. *Testing*

The proposed sensorless temperature monitoring system has been tested by simulation using Arduino Uno board and SIM900 GSM Shield. In this section, the test results are shown in Arduino serial monitor to understand the working procedure of the system background. Figure 6 shows the temperature detection, incoming SMS valid data extraction and test reliability results.

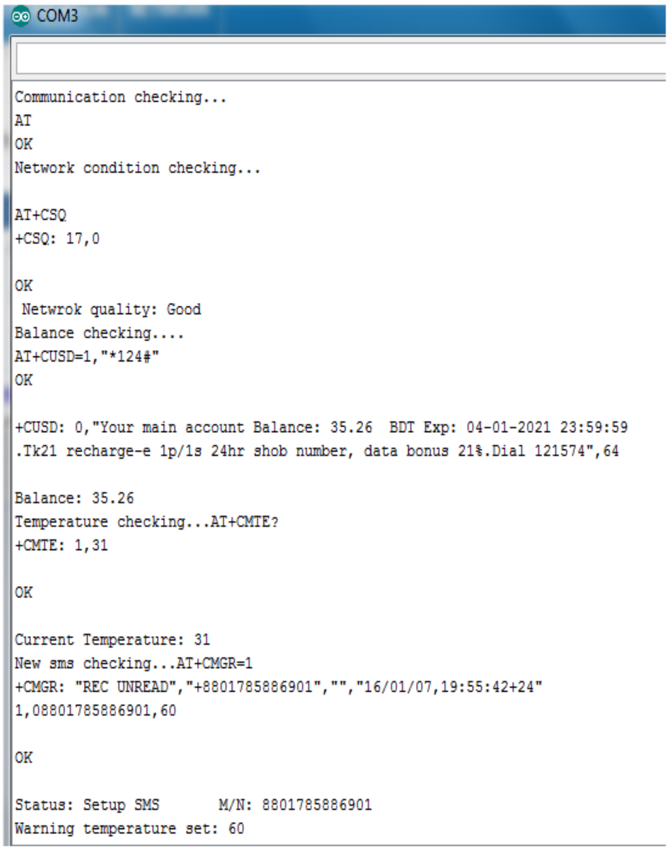


Fig. 6. AT command using and its responding data extraction

B. *Prototyping*

In the prototyping of the proposed system, a 20 by 4 Liquid Crystal Display (LCD) has been incorporated with an Arduino UNO and a SIM900 GSM shield instead of graphic LCD. A

12V 1amp power adapter was also used in the proposed system prototyping. Figure 7 shows the system frontend results displaying current temperature, action temperature, warning temperature and system's status, while all the backend works are in operation.

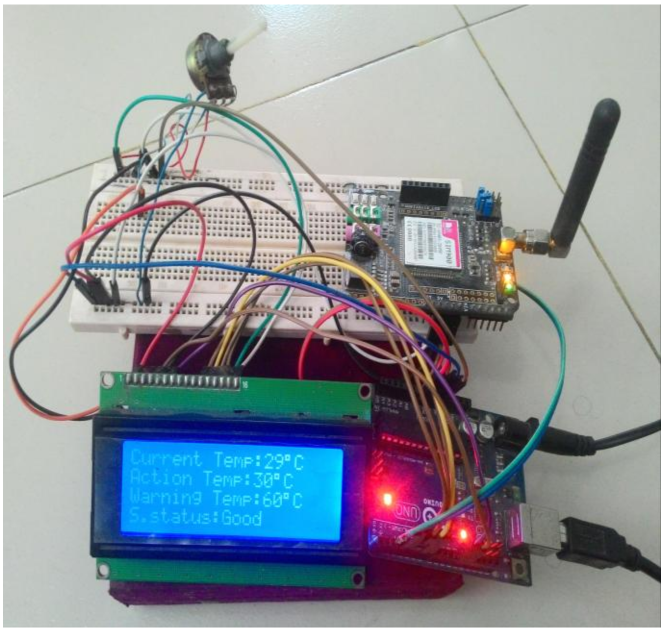


Fig. 7. Proposed system prototype

VII. CONCLUSION

A user interactive wireless smart home system has been developed which utilizes a GSM modem to access the room temperature and set the intended temperature remotely. To conclude, the temperature detection using communication chip architecture can be useful in contrast with GSM modem and external temperature sensor based smart home embedded systems since the ambient home temperature tolerance level of GSM module and external sensor varies and GSM module can be a victim of this difference.

REFERENCES

[1] Huang, Xin-Bo, Qin-dong Sun, and Han Xiao-yan. "An on-line monitoring system of temperature of conductors and fittings based on GSM SMS and Zigbee." Industrial Electronics and Applications, 2008. ICIEA 2008. 3rd IEEE Conference on. IEEE, 2008.

[2] Vaz, Alexander, et al. "Full passive UHF tag with a temperature sensor suitable for human body temperature monitoring." Circuits and Systems II: Express Briefs, IEEE Transactions on 57.2 (2010): 95-99.

[3] Liu, Hui, et al. "Development of farmland soil moisture and temperature monitoring system based on wireless sensor network." Journal of Jilin University (Engineering and Technology Edition) 38.3 (2008): 604-608.

[4] Zhang, Junguo, et al. "Forest fire detection system based on a ZigBee wireless sensor network." Frontiers of Forestry in China 3.3 (2008): 369-374.

[5] Murugan, T., AzhaPeriasamy, and S. Muruganand. "Embedded Based Industrial temperature monitoring system using GSM." International Journal of computer application 58.19 (2012).

[6] Liu, Cuihong, et al. "The application of soil temperature measurement by LM35 temperature sensors." Electronic and Mechanical Engineering and Information Technology (EMEIT), 2011 International Conference on. Vol. 4. IEEE, 2011.

[7] Ding, Jifeng, Jiyin Zhao, and Biao Ma. "Remote monitoring system of temperature and humidity based on GSM." Image and Signal Processing, 2009. CISP'09. 2nd International Congress on. IEEE, 2009.

[8] Siegmund M. Redl, Matthias K. Weber and Malcolm W.oliphant, *GSM and Personal Communication Handbook*, 1st ed., 1998.

[9] M. A. Razzak and Md. Mahamudul Hasan, "An Interactive Academic Office Display System Designed for University Based on Subscriber Identity Module Card" 5th International Conference on Electronics Engineering & Computer Science (IEMCON 2014), India.

[10] https://en.wikipedia.org/wiki/Hayes_command_set.

[11] https://en.wikipedia.org/wiki/USSD_Gateway.

[12] Typical at commands datasheet for Simcom900, http://simcom.ee/documents/SIM900/SIM900_AT%20Command%20Manual_V1.11.pdf, [18 December 2015].

Phishing Websites Detection Using Machine Learning Based Classification Techniques

Mazharul Islam
Department of Computer Science & Engineering
International Islamic University of Chittagong
Chittagong, Bangladesh
bappychw@gmail.com

Nihad Karim Chowdhury
Department of Computer Science & Engineering
University of Chittagong
Chittagong, Bangladesh
nihadcuet@gmail.com

***Abstract*— Phishing sites which expects to take the victims confidential data by diverting them to surf a fake website page that resembles a honest to goodness one is another type of criminal acts through the internet and its one of the especially concerns toward numerous areas including e-managing an account and retailing. Phishing site detection is truly an unpredictable and element issue including numerous components and criteria that are not stable. On account of the last and in addition ambiguities in arranging sites because of the intelligent procedures programmers are utilizing, some keen proactive strategies can be helpful and powerful tools can be utilized, for example, fuzzy, neural system and data mining methods can be a successful mechanism in distinguishing phishing sites. We applied different types of machine learning based classification algorithms, including Naïve Bayes (NB), Support Vector Machine (SVM), Neural Net (NN), Random Forest (RF), IBK lazy classifier and Decision Tree (J48). Finally we measured and compared the performance of the classifier in terms of accuracy.**

Keywords- Phishing Websites, Data Mining, Machine Learning, Support Vector Machine, Random Forest, Naïve Bayes, Neural Net, Decision Tree, IBK lazy classifier, WEKA.

I. INTRODUCTION

Phishing is a type of extensive fraud that happens when a malicious website act like a real one keeping in mind that the end goal to obtain touchy data, for example, passwords, account points of interest, or MasterCard numbers.

In spite of the fact that there are a few contrary to phishing programming and methods for distinguishing potential phishing endeavors in messages and identifying phishing substance on sites, phishers think of new and half breed strategies to go around the accessible programming and systems.

Phishing is a trickery system that uses a blend of social designing what's more, innovation to assemble delicate and individual data, for example, passwords and charge card subtle elements by taking on the appearance of a dependable individual or business in an electronic correspondence. Phishing makes utilization of spoof messages that are made to look valid and implied to be originating from honest to goodness sources like money related foundations, ecommerce destinations and so forth, to draw clients to visit fake sites through joins gave in the phishing email. The misleading sites are intended to emulate the look of a genuine organization site page.

The employing so as to phishing invader's trap clients diverse social building strategies, for example, debilitating to suspend client accounts on the off chance that they don't finish the account upgrade process, give other data to approve their records or a few different motivations to get the clients to visit their satirize page.

Delicate Computing strategies are progressively being utilized to address an extent of computational issues. Clustering is a kind of unsupervised learning; unsupervised learning except that there is no previous information about the class participation of the perceptions, i.e., class names of information is obscure. The reason for utilizing unsupervised learning is to specifically separate structure from a dataset without earlier preparing. On the other hand, supervised learning accommodates a vastly improved precision, unsupervised learning accommodates a quick and dependable way to deal with infer information from a dataset. That's why we used supervised learning in our work.

The remainder of this paper is organized as follows: In the next section, we provide a brief overview of the related work. In Section III, we discuss about methodology. Our experimental results are shown in Section IV. Finally, we provide the conclusion of this research in Section V.

II. RELATED WORK

Phishing site is one of the late worries in the security area. Yet, because of prominent effect on the money related and on-line retailing areas and since recognizing such sort of dangers is key towards safe web surfing, various distinctive a few promising studies and methodologies were led and proposed to this issue established in the writing. Although a considerable amount of hostile to phishing arrangements are accessible these days yet a large portion of them are not skilled to settle on a sufficiently precise choice and thus, the false-positive choices raised seriously. In this segment, we quickly depict the existed endeavors in this space through reviewing the basic related methodologies.

Nawafleh and Hadi[1] proposed new associative classification algorithm to recognizing phishing site. Observational study result demonstrates that acquainted classification is promising technique and indicated competitive execution when contrasted and different calculations, for example, SVM, PRISM, RIPPER and NB.

In [2] the study compared few learning approaches including Support-Vector-Machine, decision-trees, rule-based techniques and Bayephphishingechniques in recognizing phishing emails. A random forest algorithm was executed in PILFER (Phishing Identification by Learning on Features of Email Received) which succeeded in effectively identifying 96% of the phishing messages with a false-positive rate of 0.1%. Ten email's elements showed are utilized as a part of the experimental results those are IP address URLs, Age of Domain, Non-coordinating URLs, "Here" Link, HTML messages, Number of Links, Number of Domains, Number of Dots, Containing Java script, Spam-channel Output.

With respect to phishing location, A. Bergholz et el. [3] exhibited a methodology for enhancing learning models for recognizing phishing messages by feature selection. A subset of components is chosen by a wrapper technique in which the purported best-first pursuit calculation efficiently adds and subtracts features to a present subset utilizing the classifier itself as a feature of the evaluation function.

Pradeep and Ravendra [4] proposed a model that can detect a website is phishing or not. They uses six different machine learning based classification algorithm named Naïve Bayes, J48, SVM, Random Forrest, Tree Bag and IBK lazy classifier with 92.7846%, 95.11%, 96.57%, 96.3%, 93.85%, 93.4039% classification accuracy respectively. Their split ratio is 70-30. Where 70% accounts to training and rest for testing. Our aim is to extend their work to gain more classification accuracy using those algorithms and also we introduced a new classification algorithm named Neural Net in this experiment.

III. METHODOLOGY

A. Training and Classification

The common approach for classification problems is supervised learning. That's why in this paper we used different supervised machine learning algorithm to get the desired output for detecting the phishing websites properly. In next few paragraphs we have discussed about different supervised learning algorithm which we used in the experiment.

- Naïve Bayes

  This classification algorithm based on applying Bayes theorem which is also known as a probabilistic algorithm [6]. This algorithm used in classification because of its simplicity in both during training and classifying stage. Another advantage of this algorithm is less data needed during training stage compared to the other's machine learning based classification algorithm.

  For a document d and class c, by Bayes theorem,

$$P(c \mid d) = \frac{P(d \mid c)P(c)}{P(d)}$$

  Then Naïve Bayes classifier will be,

$$c* = \arg\max cP(c \mid d)$$

- J48

  J48[QUI93] implements Quinlan`s C4.5 algorithm [QUI92] for creating a pruned or unpruned C4.5 decision tree. C4.5 is an augmentation of Quinlan's prior ID3 algorithm. The decision trees created by J48 can be utilized for classification. J48 constructs decision trees from an arrangement of labeled training data utilizing the idea of data entropy. It utilizes the way that every quality of the information can be utilized to settle on a choice by part the information into littler subsets. J48 looks at the standardized data pick up (distinction in entropy) that outcome from picking a trait for part the data. To settle on the choice, the property with the most noteworthy standardized data increase is utilized. At that point the calculation repeats on the littler subsets. The part technique stops if all instances in a subset have a place with the same class. At that point a leaf node is made in the choice tree advising to pick that class. In any case, it can likewise happen that none of the components give any data pick up. For this situation J48 makes a decision node higher up in the tree utilizing the normal estimation of the class. J48 can deal with both continuous and discrete attributes, preparing data with missing property estimations and qualities with varying expenses. Further it gives an alternative to pruning trees after creation.

- Support Vector Machine

  This is a well known machine learning based classification algorithm. Support Vector Machine (SVM) [7] is based on the concept of decision planes that define decision boundaries. The decision plane also known as hyper plane which separates between a set of objects that having different class memberships.

  SVM is a standout amongst the most popular classifiers nowadays. The thought here is to discover the operation optimal isolating hyperplane between two classes by maximizing the edge between the classes nearest focuses. Assume that we have a straight isolate capacity and two directly divisible classes with target values +1 and - 1. A discriminating hyperplane will satisfy:

$$w'x_i + w_0 \geq 0 \; if t_i = +1$$

$$w'x_i + w_0 < 0 \; if t_i = -1$$

Now the distance of any point x to a hyperplane is $|wx_i + w_0|/||w||$ and the distance to the origin is $|w_0|/\ ||w||$

- Neural Net

A neural system is organized as an arrangement of interconnected indistinguishable units (neurons). The interconnections are utilized to send signals from one neuron to the next. Also, the interconnections have weights to upgrade the conveyance among neurons [8]. The neurons are not capable by them-selves; in any case, when associated with others they can perform complex calculations. Weights on the interconnections are overhauled when the system is prepared, consequently significant interconnection assume more part amid the testing stage. Since interconnections do not circle back or skip different neurons, the system is called feed forward. The force of neural systems originates from the nonlinearity of the concealed neurons. In result, it is huge to acquaint nonlinearity in the system with have the capacity to learn complex mappings. The ordinarily utilized capacity as a part of neural system examination is the sigmoid capacity, which has the structure.

- Random Forest :

Random forest [9] is a classifier that joins numerous tree predictors, where every tree relies on upon the estimations of an irregular vector inspected autonomously. Besides, all trees in the forests have the same appropriation. Keeping in mind the end goal to develop a tree we expect that n is the quantity of preparing perceptions furthermore, p is the quantity of variables (elements) in a preparation set. Keeping in mi nd the end goal to decide the choice hub at a tree we pick K << p as the quantity of variables to be chosen. We select a bootstrap test from the n perceptions in the preparation set furthermore; utilize whatever is left of the perceptions to assess the blunder of the tree in the testing stage. Subsequently, we haphazardly pick k variables as a choice at a specific hub in the tree and calculate the best split in light of the k variables in the preparation set. Trees are constantly developed and never pruned contrasted with other tree calculations. Irregular backwoods can deal with expansive quantities of variables in an information set. Likewise, amid the backwoods building process they generate an inside fair-minded appraisal of the speculation mistake. What's more, they can assess missing information well. A noteworthy downside of arbitrary timberlands is the absence of reproducibility, as the procedure of building the timberland is arbitrary. Further, clarifying the final model and ensuing results is difficult, as it contains numerous free choices trees.

- IBK lazy classifier

IBK is a k-nearest neighbor classifier that uses the same distance metric. The quantity of closest neighbors can be indicated expressly in the article editorial manager or decided naturally utilizing forget one cross-approval center to a maximum utmost given by the predetermined quality. IBK is a k nearest neighbor classifier. A sort of various search algorithms can be utilized to accelerate the assignment of finding the closest neighbors. A linear search is a default yet encourages choices incorporate KD-trees, ball trees, thus called "spread trees". The separation capacity utilized is a parameter of the search technique. The remaining thing is the same with respect to IBL—that is, the Euclidean separation; different choices incorporate Chebyshev, Manhattan, and Minkowski separations [10]. Predictions from more than one neighbor can be weighted by separation from the test pattern and two distinct equations are executed for changing over the separation into a weight [11][12].

The principle advantage picked up in utilizing a lazy learning technique is that the objective capacity will be approximated locally, for example, in the k-closest neighbor algorithm. The disadvantage with lazy learning incorporates the expansive space prerequisite to store the complete training dataset.

IV. EXPERIMENTAL RESULTS

A. Dataset Description

In our experiment we use the phishing websites dataset available at Machine Learning Repository [5]. The dataset consists of 11055 websites samples. Each sample consists of 31 attributes. All of samples are already labeled with 1 and -1.

Table 1 shows the classification results of the six generated classifiers using all features from the dataset.

TABLE I. CLASSIFICATION ACCURACY

Classification Accuracy(%)	
Naïve Bays	93.09
J48	95.61
SVM	94.84
Neural Net	96.65
Random Forrest	**97.47**
IBK lazy classifier	97.07

As we can see clearly from the table that the lowest classification accuracy gained by the Naïve Bays classifiers and the highest classification accuracy is gained by Random Forest. Though the classification accuracy of IBK lazy classifier is 97.07% which is close to the Random Forest. Although Support Vector Machine and J48 shows a promising accuracy rate. We also introduced a new classification algorithm in this experiment which is Neural Net and it shows a good classification accuracy compare to others. The highest accuracy of the base paper was 96.57%, gained by the Support Vector Machine. We were able to extend the result to 97.47% by Random Forest. Also we extend their accuracy of the other classifiers. Naïve Bays from 92.8746% to 93.0941%, J48 from 95.11% to 95.5971% and IBK lazy classifiers from 93.4039% to 97.0748%.

V. CONCLUSION

Phishing is a cyber crime procedure utilizing both social building and specialized deception to take individual sensitive data. Besides, Phishing is considered as another extensive type of fraud. Experimentations against recent dependable phishing data sets utilizing different classification algorithm have been performed which received different learning methods. The base of the experiments is accuracy measure.

The aim of this research work is to predict whether a given URL is phishing website or not. It turns out in the given experiment that Random forest based classifiers are the best classifier with great classification accuracy of 97.47% for the given dataset of phishing site. As a future work we might use this model to other Phishing dataset with larger size then now and then testing the performance of those classification algorithm's in terms of classification accuracy.

REFERENCES

[1] S. Nawafleh, W. Hadi (2012). Multi-class associative classification to predicting phishing websites. International Journal of Academic Research Part A; 2012;4(6), 302-306J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[2] Sadeh N, Tomasic A, Fette I. Learning to detect phishing emails. Proceedings of the 16th international conference on World Wide Web. 2007: p. 649-656.

[3] Andr Bergholz, Gerhard Paa, Frank Reichartz, Siehyun Strobel, and Schlo Birlinghoven. Improved phishing detection usingmodel-based features. In Fifth Conference on Email and Anti-Spam, CEAS, 2008

[4] P. Tiwari, R. Singh International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181Vol. 4 Issue 12, December-2015.

[5] UCI Machine Learning Repository." http://archive.ics.uci.edu/ml/, 2012.

[6] H. A. Chipman, E. I. George, and R. E. McCulloch.BART: Bayesian Additive Regression Trees. Journal of the Royal Statistical Society, 2006. Ser.B,Revised.

[7] J. P. Marques de Sa. Pattern Recognition: Concepts,Methods and Applications. Springer, 2001.

[8] D. Michie, D. J. Spiegelhalter, and C. C. Taylor. Machine Learning, Neural and StatisticalClassi_cation. Ellis Horwood, 1994.

[9] L.Breiman.Random forests. Machine Learning,45(1):5{32, October2001

[10] Mrs. Sayantani Ghosh, Mr. Sudipta Roy, Prof. Samir K.Bandyopadhyay, "A tutorial review on Text Mining Algorithms".

[11] Kaushik H. Raviya, Biren Gajjar, "Performance Evaluation ofDifferent Data Mining Classification Algorithm Using WEKA".

[12] Dr. S.Vijayarani, S.Sudha,"Comparative Analysis of Classification Function Techniques for Heart Disease Prediction".

Performance Analysis between Probabilistic and Decision Tree based Classification on User Knowledge Model Dataset

Md. Faisal Alam[1], Rashik Hafiz[1], Mohammad Sanaullah Chowdhury[2], Nihad Karim Chowdhury[2]
[1]Department of Computer Science And Engineering,
International Islamic University Chittagong, Bangladesh
Email: md.alamfaisal@gmail.com, rashikhfz@gmail.com
[2]Department of Computer Science And Engineering,
University of Chittagong, Bangladesh
Email: mohammadsanaullahchowdhury@gmail.com, nihadcuet@gmail.com

***Abstract*—Classification is an important data mining technique to classify various kinds of data used in our daily life. It is used to classify the item according to the features of item with respect to the predefined set of classes. Probabilistic classification algorithm Naïve Bayes and Decision Tree based J48 algorithm generate superior result in various problem domain. This paper is a summary on performance evolution based on correct instances of data using these classification algorithms. Here, we use UCI user knowledge modeling dataset to analyze results. It is a challenge to determine efficiently what type of data will be stored and how it will be implemented by a knowledge modeling system. From experimental evaluation, it is revealed that the efficiency and accuracy of J48 is better than Naïve Bayes.**

***Index Terms*—Classification, Naïve Bayes, Decision Tree J48, UCI User knowledge Modeling Dataset**

I. INTRODUCTION

Recently, use of machine learning techniques in data mining is developing. Data mining is a knowledge base subfield of computing that is growing in various applications widely. It is an analytical process designed to explore patterns in massive information sets involving methods at the intersection of computer science, machine learning, statistics, and information systems. Different methodologies of data mining are mainly classified as classification, clustering, association and regression. Classification analysis is the organization of data in given classes. It is also known as Supervised Classification[2]. Supervised Classification is a method which make a function from training data. In this paper, we focus on the data classification and measure performance of the classifier algorithm Naïve Bayes and J48.

Naïve Bayes classifier uses supervised learning methodology for classification [1]. It is based on application of Bayes theorem with independent probabilities from every pair of features. Also used as a baseline classifier that give reasonable classification performance[12]. Another one, J48 is the implementation of algorithm Iterative Dichotomiser 3 which is developed by the WEKA project team. J48 is using decision tree for classification by creating a binary tree. Decision trees are generated by using greedy technique and it uses reduced error pruning[2].These algorithms are so effective for classification.

To measure performance of these algorithms, we use UCI user knowledge data set. User information modeling system is one among the foremost powerful mechanisms for the web-based accommodative applications. The goal is to provide enough or knowledge for students/users. User knowledge modeling is increasing the classification accuracy of the well-known and usually used probabilistic and instance-based user modeling algorithms[3]. It is additionally presenting a practicable and simply to know an intuitive data classifier algorithms producing effective results to the user modeling literature.

The remainder of this paper is organized as follows: In the next section, we provide a brief overview of the related work. In Section III, we discuss about methodology. Our experimental results are shown in Section IV. Finally, we provide the conclusion of this research in Section V.

II. RELATED WORK

Many works associated with the comparison of the classification strategies are done. One task of the user modeling systems is to trace and store user actions. Another task is to judge and convert the keep data into useful information. Within the analysis method of user information, mostly rule-based approaches and machine learning algorithms particularly for classification, prediction and clustering are used [8]. OMahony and Smyth [9] with Naïve success described a supervised classification process to design to spot and suggest the most useful product reviews. They compared the performance efficiency of JRip, J48 and Naïve Bayes classification techniques using a vary of options derived from building reviews. Reviewing samples consisted of options derived from four distinct classes well-mined from individual reviews and also the wider community reviewing activity. Prior to classification, every review was translated into a feature-based instance representation. The disadvantage of this approach is that all of the options have same impact in the decision-making method of classifiers. The experimental studies, the AUROC scores (area under mythical monster curve) of classifiers failed to exceed 82%. Weight of the options can be increasing the scores of AUROC classifiers.

Huang et al.[10] compared Naïve Bayes, decision Tree and Support Vector Machine with each other mistreatment area under Curve criterion. Area under Curve criterion is healthier than accuracy for comparing the classification strategies [11]. Moreover, it's shown that C4.5 implementation of Decision Tree has higher area under Curve compared to Naïve Bayes and Support Vector Machine. Yogendra Kumar Jain [13], compared J48, BayesNet, OneR and, NB algorithms for intrusion detection which shows that the J48 decision tree gives more accuracy. In that year, Rangaduari [14] introduces a Adaptive NIDS using a Hybrid approach alone with two stages. In the first stage a probabilistic classifier is used then a HMM based traffic model is used in the second stage. Kahraman et al.[3] state that to beat the problems of knowledge classifier, an Intuitive k-NN knowledge Classifier (IKC) has been introduced, The projected IKC technique achieved a mean classification accuracy of 97.5% over the validation set, whereas the classification accuracy of the opposite approaches ranged from 73.8% to 85%. Though, IKC approach has created a substantial improvement.

III. METHODOLOGY

Classification approaches usually use a training set where all objects are already related to noted class labels[3]. The classification algorithm creates a model by using training dataset. The model is used to classify new objects. In this paper, Naïve Bayes algorithm and J48 decision tree are used for comparison. Comparison is made on accuracy, sensitivity and specificity using testing instances. Correct instances that give us suggestion a most efficient method for classification.

A. Naïve Bayes Classifier:

The Naïve Bayes algorithm is a straightforward probabilistic classifier that calculates a set of possibilities by investigation the frequency and combinations of values in a given information set[4]. The rule uses Bayes theorem which assumes all attributes to be independent where the values are getting from class variable. This conditional independence aspect rarely holds true in real world applications. Therefore the characteristics of this algorithm tend to perform well and learn quickly in various supervised classification problems. The main advantage of this classifier is that, it only needs a small amount of training data to verify the parameters needed for classification. Conditional probability of Naïve Bayes classifier as given below.

$$P\left(C_n \mid x\right)$$

Here, C_n is the n^{th} class which is conditional on a input vector named x. Where,

$$x = \sum_{j=1}^{n} x_j$$

Using Naïve Bayes theorem, the equation become.

$$P\left(C_n \mid x\right) = \frac{P\left(x \mid C_n\right) P\left(C_n\right)}{P\left(x\right)}$$

In the above equation, $P(C_n|x)$ is the posteriori probability of C conditioned on x. $P(C_n)$ and P(x) is the priori probability of C and x continuously wherever $P(C_n)$ is independent of x. Again, $P(x|C_n)$ is the posteriori probability of x conditioned on C.

B. Decision Tree J48:

Decision tree is a classification technique, Based on divide and conquers strategy. A decision tree consist decision nodes and leaf nodes, where decision node specifies a test over one of the attributes and a leaf node represents the class value. Every path from the root node to leaf node is rule. Classification error is the performance major factor for Decision tree. Classification error is defined as the percentage of misclassified cases. J48 is an open source Java implementation format of the C4.5 decision tree algorithm for classification. It creates a binary tree, which built to model the classification process[5]. When the tree is built, it is applied to every tuple within the information and ends up in classification for that tuple. While building a tree, J48 ignores the missing values. Value for that item will be expected based on what's identified about the attribute values for the other records[6]. The main plan is to divide the data into many ranges which based on the attribute values for that item whose are found in the training set[7]. It allows classification via decision trees or rules generated from them. Algorithm 1 shows the procedure of Decision Tree J48.

Algorithm 1 Decision Tree J48

```
N ← rootNode
if T belongs to same category C then
    leafNode ← N
    mark N as a class C
    return C
end if
for i ← 1 to N do
    Calculate informationGain of A_i
end for
t_x ← testingAttribute
highestInformationGain ← attribute which have Highest informationGain
N.t_x ← highestInformationGain
if N.t_x is Continuous Process then
    find thresholdValue
end if
for all TSplit in the Splitting of T do
    if TSplit is Empty then
        Child of N ← leafNode
    else
        Child of N ← dTree(TSplit)
    end if
end for
Calculate Error Rate of Node N
Return N
```

IV. RESULT

We use Weka[15] tool which stands for Waikato Environment for Knowledge Analysis for classification using Naïve Bayes and J48 decision tree algorithm on UCI user knowledge model dataset. These algorithms are already built in Weka. We classify dataset using cross-validation and percentage split feature of Weka. To do cross validation, we divide Dataset into 10, 20, 30, 40, and 50 part using folds option. We gave percentage rates 50, 60, 70, 80, 90 and 92 on percentage split option to train the classifier.

A. Dataset:

We use UCI User Knowledge Modeling Data Set (https://archive.ics.uci.edu/ml/datasets/User+Knowledge+Modeling). It is real dataset about the students knowledge status about the subject of Electrical DC Machines. Undergraduate students of Department of Electrical Education of Gazi University in the 2009 semester denoted this data.The sample dataset is following. Fig 1 shows example for data instances.

1	STG	SCG	STR	LPR	PEG	UNS
2	0	0	0	0	0	very_low
3	0.08	0.08	0.1	0.24	0.9	High
4	0.06	0.06	0.05	0.25	0.33	Low
5	0.1	0.1	0.15	0.65	0.3	Middle
6	0.08	0.08	0.08	0.98	0.24	Low
7	0.09	0.15	0.4	0.1	0.66	Middle
8	0.1	0.1	0.43	0.29	0.56	Middle
9	0.15	0.02	0.34	0.4	0.01	very_low
10	0.2	0.14	0.35	0.72	0.25	Low

Fig. 1: Sample Dataset

Attribute Information:

Attributes are set by the dataset creator.
STG - The degree of study time for goal object materials.
SCG - The degree of repetition number of user for goal object materials.
STR - The degree of study time of user for related objects with goal object.
LPR - The exam performance of user for related objects with goal object.
PEG - The exam performance of user for goal objects.
UNS - The knowledge level of user.

The users' knowledge class was classified by the authors.

Using Naïve Bayes Classifier on Weka, Table I and Table II shows the summary of Cross and Split Validation respectively. And Table III and Table IV shows the summary of Cross and Split Validation respectively by using J48 Decision Tree.

The Tables I, II, III and IV show the results according to total 258 instances. In Naïve Bayes, percentage is max when the folding value is 10 and 20 in cross-validation. Applying J48 Decision Tree algorithm, we get the better result whenever fold value is become 20. Although, the percentage result is

TABLE I: Cross Validation for Naïve Bayes

Fold	Classified Instance	Number of Instance	Percentage
10	Correct Incorrect	230 28	**89.1473** 10.8527
20	Correct Incorrect	230 28	**89.1473** 10.8527
30	Correct Incorrect	229 29	88.7597 11.2403
40	Correct Incorrect	229 29	88.7597 11.2403
50	Correct Incorrect	229 29	88.7597 11.2403

TABLE II: Split Validation for Naïve Bayes

Split %	Classified Instance	Number of Instance	Percentage
50	Correct Incorrect	109 20	84.4961 15.5039
60	Correct Incorrect	88 15	85.4369 14.5631
70	Correct Incorrect	64 13	83.1169 16.8831
80	Correct Incorrect	44 8	84.6154 15.3846
90	Correct Incorrect	22 4	84.6154 15.3846
92	Correct Incorrect	18 3	**85.7143** 14.2857

TABLE III: Cross Validation for J48

Fold	Classified Instance	Number of Instance	Percentage
10	Correct Incorrect	241 17	93.4109 6.5891
20	Correct Incorrect	242 16	**93.7984** 6.2016
30	Correct Incorrect	238 20	92.2481 7.7519
40	Correct Incorrect	241 17	93.4109 6.5891
50	Correct Incorrect	241 17	93.4109 6.5891

TABLE IV: Split Validation for J48

Split %	Classified Instance	Number of Instance	Percentage
50	Correct Incorrect	13 16	87.5969 12.4031
60	Correct Incorrect	96 7	93.2039 6.7961
70	Correct Incorrect	68 9	88.3117 11.6883
80	Correct Incorrect	49 3	94.2308 5.7692
90	Correct Incorrect	25 1	96.1538 3.8462
92	Correct Incorrect	21 0	**100** 0

TABLE V: Maximized Result Summary

Method	Cross Validation	Percentage
	20	92%
Naive Bayes	89.1473%	85.7143%
J48	**93.7984%**	**100%**

maximum at the splitting value become 92% to train these classifier. Analyzing the results we get that J48 Decision Tree algorithms performance is efficient and best.The efficiency become 100% by applying J48 Decision Tree. Table V shows the Summary of experiment when the algorithms performance result is max.

Overall we state that J48 decision tree algorithm is better than Naïve Bayes classifier algorithm on this dataset.

V. CONCLUSTION

In this research, we focus on performance comparison between Naïve Bayes and J48 algorithm on User Knowledge Model dataset. Experimental results shown in the study are about classification accuracy of both algorithms. Here we get, Decision Tree J48 algorithm is better than Naïve Bayes classifier. From experimental evaluation it is also revealed that as the number of cross validation fold is increased, classification accuracy for both Naïve Bayes and J48 algorithm is also increased. We get 100% accuracy while using J48 Decision Tree algorithm. For any knowledge base modeling systems classifier J48 is perfect to classify classes.Furthermore, The Performance of a student can be measured by this algorithm and suggest students what to do next.

REFERENCES

[1] Bhosale, Dipali, Roshani Ade, and P. R Deshmukh. ”Feature Selection based Classification using Naïve Bayes, J48 and Support Vector Machine.” International Journal of Computer Applications 99.16 (2014): 14-18

[2] Patil, Tina R., and S. S. Sherekar. ”Performance analysis of Naïve Bayes and J48 classification algorithm for data classification.” International Journal of Computer Science and Applications 6.2 (2013): 256-261.

[3] Kahraman, H. Tolga, Seref Sagiroglu, and Ilhami Colak. ”The development of intuitive knowledge classifier and the modeling of domain dependent data.” Knowledge-Based Systems 37 (2013): 283-295.

[4] El-Gohary, Hatem, ed. Transdisciplinary Marketing Concepts and Emergent Methods for Virtual Environments. IGI Global, 2012 : page 211.

[5] Pereira, Francisco, et al., eds. Progress in Artificial Intelligence: 17th Portuguese Conference on Artificial Intelligence, EPIA 2015, Coimbra, Portugal, September 8-11, 2015. Proceedings. Vol. 9273. Springer, 2015.

[6] Abraham, Ajith, Pavel Krmer, and Vaclav Snasel, eds. Afro-European Conference for Industrial Advancement: Proceedings of the First International Afro-European Conference for Industrial Advancement AECIA 2014. Vol. 334: page 184, Springer, 2014.

[7] Dunham, Margaret H. Data mining: Introductory and advanced topics. Pearson Education India, 2006:page 96.

[8] Pooja Sharma, Annu Mishra, Classification Algorithm Using Random Concept On A Very Large Data Set: A Survey , IJMTER-2014

[9] I. Colak, S. Sagiroglu, H.T. Kahraman, A user modeling approach to web based adaptive educational hypermedia systems, Seventh International Conference on Machine Learning and Applications, IEEE Computer Society, 2008, pp. 694699.

[10] M.P. OMahony, B. Smyth, A classification-based review recommender, Knowledge-Based Systems 23 (2010):323329.

[11] J. Huang, J. Lu, C.X. Ling, Comparing Naïvee Bayes, Decision Trees, and SVM with AUC and Accuracy, Proceedings of the Third IEEE International Conference on Data Mining, 2003.

[12] Vaidya, Jaideep, Murat Kantarcolu, and Chris Clifton. ”Privacy-preserving naive bayes classification.” The VLDB JournalThe International Journal on Very Large Data Bases 17.4 (2008): 879-898.

[13] Yogendra Kumar Jain and Upendra, An Efficient Intrusion Detection Based on Decision Tree Classifier Using Feature Reduction, International Journal of Scientific and Research Publications, vol. 2, issue 1, ISSN 2250-3153, Jan. 2012

[14] Rangadurai Karthick R., Hattiwale V.P., and Ravindran B., Adaptive network intrusion detection system using a hybrid approach, 4th International Conference on Communication Systems and Networks (COMSNETS), vol.1, no. 7, pp. 3-7, Jan. 2012

[15] Hall, Mark, et al. ”The WEKA data mining software: an update.” ACM SIGKDD explorations newsletter 11.1 (2009): 10-18.

Novel Approach of Phased Array Antenna with Beam Steering Technology for Microwave Power Transmission from SSPS System

Golap Kanti Dey [1], Kazi Tanvir Ahmmed [2], Rubell Sen Goopta [1]

[1] School of Engineering and Computer Science, Chittagong Independent University, Chittagong-4000, Bangladesh.

[2] Department of Applied Physics, Electronics & Communication Engineering, University of Chittagong, Chittagong-4331, Bangladesh

Email: golap@ciu.edu.bd, tanvir@cu.ac.bd, rubellsen@ciu.edu.bd

Abstract- **This paper represents design methodology of novel phased array antenna and its radiation pattern of 100 elements linear phased array for highly efficient microwave power transmission (MPT) which will be focused on ground based receiving station for microwave power transmission from Space Solar Power Satellite (SSPS) system. Taking different ratio between antenna element spacing, d and wavelength, λ we studied radiation pattern of phased array antenna for proposed 2.45 GHz microwave power input. Phased array antenna is a multiple-antenna system, the prominent part of SSPS system, in which the radiation pattern can be reinforced in a particular direction and suppressed in undesired directions by steering electronically in the microwave power transmission from SSPS system.**

Keywords— Phased Array Antenna, Array Construction, N-element phased array, Radiation Pattern etc.

I. INTRODUCTION

Antenna is a dominant technology required for Solar Power Satellite system. Solar power satellite (SPS) [1] is a renewable and infinite energy system which works in the Geostationary Earth Orbit as an electric power plant in space. Main theme of solar power satellite is that SPS will collect solar energy, converts sunlight to electricity and beam the power to ground-based receiving stations. Among the four main parts of microwave power transmission (MPT) for SSPS system, antenna is prominent part where we are proposing the noble phased array antenna (PAA) for reinforcing the radiation pattern precisely and accurately.

All kinds of antennas can be applied as a radiating part for MPT system, as for example, parabolic antenna, microstrip antenna, Yagi-Uda antenna, horn antenna etc. However to control a microwave beam direction accurately and precisely, we have to use a phased array antenna system.

Phased array antennas [2] consist of multiple fixed antenna elements to reinforce the radiation pattern in a specific direction with suppressing the undesired one. Relative phase, amplitude applied to each radiating element are used to determine the shape and direction of radiation pattern. By using the phased array antenna we can steer the beam pattern electronically. By controlling the amplitude and phase of each element individually beam pattern of the array can be formed to be more generalized. Beamforming by using this technique can be used to suppress side lobes, by creating the patterns of radiation to the specific direction.

From the several decades phased arrays have been traditionally used in military applications. . Increasing interest has drawn in utilizing phased array technology for commercial applications after recent growth in civilian radar-based sensors and communication systems. Hsi-Tseng Chau [3] designed a methodology of a phased array antenna whose radiation will focus in the near zone of array aperture by using microstrip feeding circuits. Takenori Yasuzumi [4] worked on new type of phased array antenna using bi-layered microstrip antenna (MSA) composed by 3-patch element. Reference [5] proposed a measurement method that can reduce measurement time for phased array antenna while providing all radiation patterns and a fixture to measure the 3-D radiation pattern which is compact and provides low interference on antenna's radiation pattern. Naoki Shinohara [6] did their experiment for Microwave Power Transmission with an advanced phased array system to make experiments on beam forming with phased array. Finite difference time-domain method to the generalized analysis of phased array antennas [7] also presented by Gregory M. Turner. Low-profile. High-sensitivity receiving sub-array module [8] with high-temperature superconducting filters for an active phased array antenna also developed by Hiroyuki Kayano. Latest research work has been going on to use phased antenna in the wireless power transfer. In this research work we are going to depict phased array antenna system to transfer microwave power which is the significant part of SSPS system.

In our research work we are presenting the uniform linear phased array system with the radiation pattern for 100 elements where proposed microwave power input is 2.45 GHz for SSPS system. We have also depicted radiation patterns by varying ration between the distance of the element and wavelength. Linear arrays are designed to produce a narrow beam of main lobe keeping side lobes small as far as possible. The primary reason for using phased array is to produce a directive beam that can be repositioned (scanned) electronically.

II. PHASED ARRAY CONSTRUCTION

Phased array antenna is a directive antenna system consist of multiple fixed antenna elements with individual radiating sources feeding coherently and use alterative phase or time-delay control at each element to reinforce the radiation pattern to predefined angles in a particular direction. For pattern shaping sometimes variable amplitude control is also provided. However, the fundamental reason for using phased arrays is to generate beam pattern that can be steered [9] by means of electronic control system.

Block diagram of an N-element phased array with variable delay element is shown in Fig.1 where "N" identical antennas are uniformly spaced by a distance "d" along an axis.

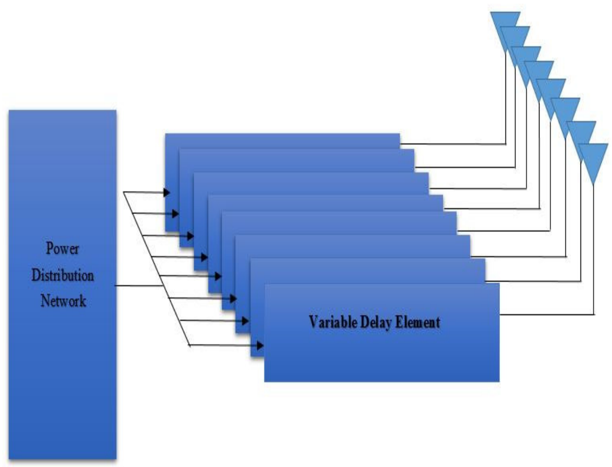


Fig. 1- Block diagram of N-element phased array

Individual variable time delay elements are incorporated at each signal path before transmitting the signals to control the phases of the signals. At an angle of θ to the normal direction a plane-wave beam is assumed to be incident upon the antenna array. Due to the spacing between the antenna elements, the radiating element will experience a time delay [10], as written in Eqn.1, to reach the successive antennas.

$$\Delta\tau = \frac{2\pi d \sin(\theta)}{\lambda} \tag{1}$$

Here, λ is the wavelength of the signals.

However, signals received by each of the antennas can be written as Eqn. 2 after assuming incoming signal is a sinusoid at frequency ω with amplitude of A.

$$S_i = Ae^{-jn\Delta\tau} \tag{2}$$

To compensate linear delay progression of the signal arrived at the successive phased array antenna elements due to the spacing between the elements, variable delay circuits must be introduce analogous but with reverse delay progression. In the uniform linear arrays, variable time delays are designed to allow uniform phase progression across the array. Hence, the output signal in each channel of the variable delay block can be written as Eqn. 3

$$S'i = Ae^{-jn\Delta\tau} e^{-jn\alpha} \tag{3}$$

Where α stands for the difference in phase shift provided by two successive variable time delay elements.

However, array factor [10] equal to the sum of all the signals normalized to the signal at one path can be written as Eqn. 4.

$$F = \sum_{n=1}^{N} e^{-jn(\Delta\tau - \alpha)} \tag{4}$$

According to Eqn.4, the peak of the array factor occurs at an incident angle which can be determined by Eqn. 5.

$$\frac{2\pi d}{\lambda}\sin(\theta) = \alpha \tag{5}$$

At this incident angle, which is known as scan angle, the linear delay progression introduced by the wave arriving at the successive antennas is accurately compensated with the time delay elements incorporated at each path. The array factor can also be shown as Eqn. 6.

$$F = \frac{\sin^2[\frac{N}{2}(\frac{2\pi d}{\lambda}\sin(\theta_{in}) - \alpha)]}{\sin^2[(\frac{2\pi d}{2\lambda}\sin(\theta_{in}) - \alpha)]} \tag{6}$$

The array factor has a maximum value of N^2 [10] at the scan angle "θ". F will be lower than this value indicating spatial selectivity of phased array for other angles of incident. One of the key capabilities of PAA is that instead of using mechanical rotation of antenna array we can increase the peak gain of the array through exercising electronically tunable variable time delay elements.

By increasing the number of phased array elements we can increase the efficiency of the phased array antenna. In addition, to enhance the spatial selectivity of phased array the beam width of the array can be reduced by enhancing the number of array elements.

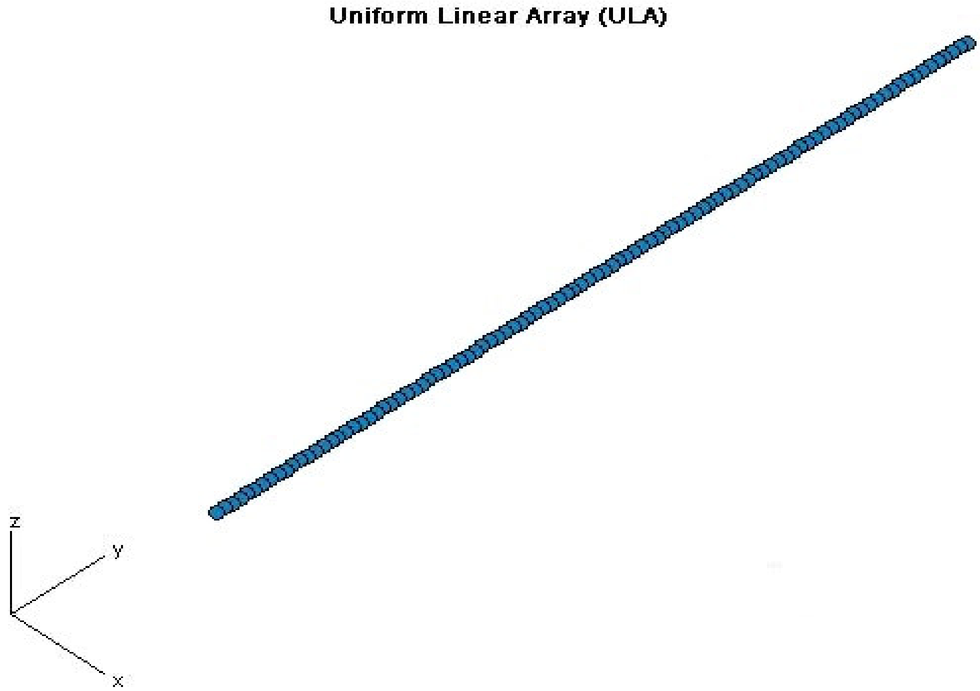


Fig.2. Uniform Linear Array

III. UNIFROM LINEAR ARRAY

We can make the structure of the linear phased array [11] [12] arranging the elements in a straight line in one dimension. These antennas within a straight line in one dimension are fed about a common phase shifter or time delay elements. Linear arrays are designed to produce a narrow beam width. Linear antenna arrays can have uniform or non-uniform spacing between elements. Commonly used linear antenna array is the Uniform Linear Array shown in Fig.2 where we have simulated ULA for 100 elements keeping element spacing 100mm with aperture size of 10m in Y-axis.

IV. RADIATION PATTERNS

A common notation in the antenna literature is used here, where λ is wavelength, d is element spacing .Now we will show how the radiation pattern changes as the parameters are modified. We have simulated all the radiation pattern by using MATLAB for the 100 elements phased array varying the ratio between the uniform distance of the phased array element and wavelength of the signal.

For the ratio between the uniform distance of the phased array element and wavelength of the signal, d/ λ =0.05 we get the simulated beam pattern in Fig.3 with a wide main lobe with several side lobes.

For SSPS we need narrower main lobe at the center by suppressing the side lobes as far as possible.

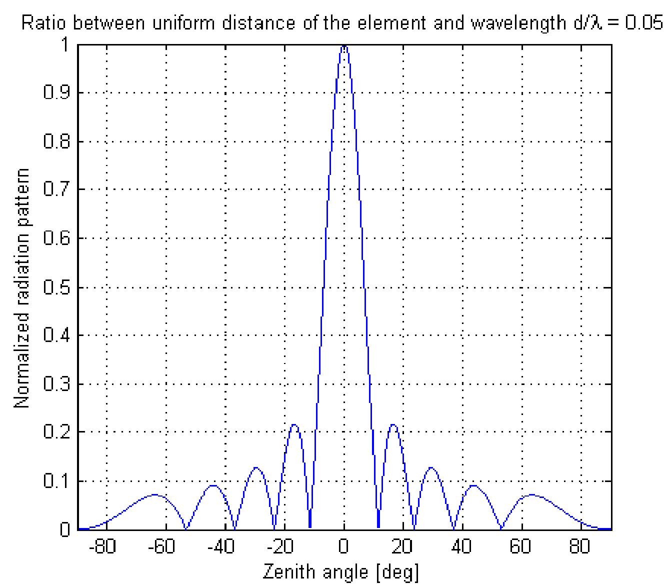


Fig.3 Radiation pattern when ratio between distance and wavelength d/λ=0.05

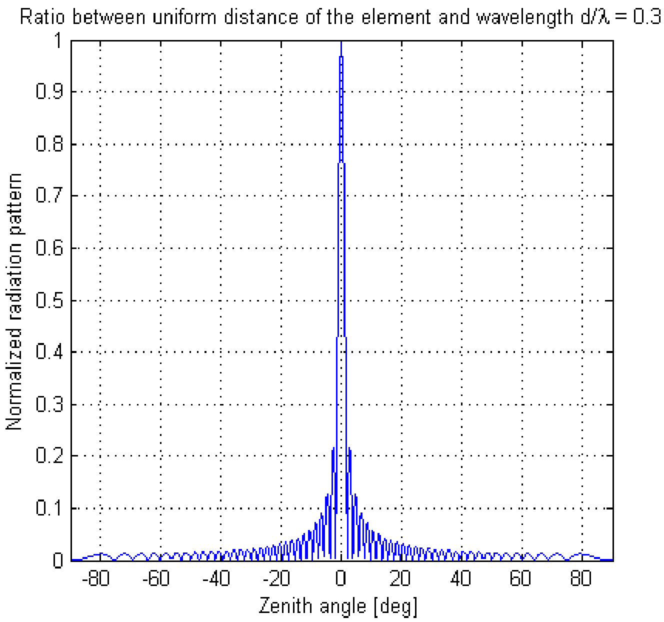


Fig.4 Radiation pattern when ratio between distance and wavelength d/ λ =0.3

After taking the ratio between the uniform distance of the phased array element and wavelength of the signal, d/λ =0.3 shown in Fig.4 we get narrower main lobe with some small side lobes. After increasing the ratio of d/λ from 0.05 to 0.3 main beam width seems to be narrower than the earlier one with minimizing and decreasing the height of the side lobes.

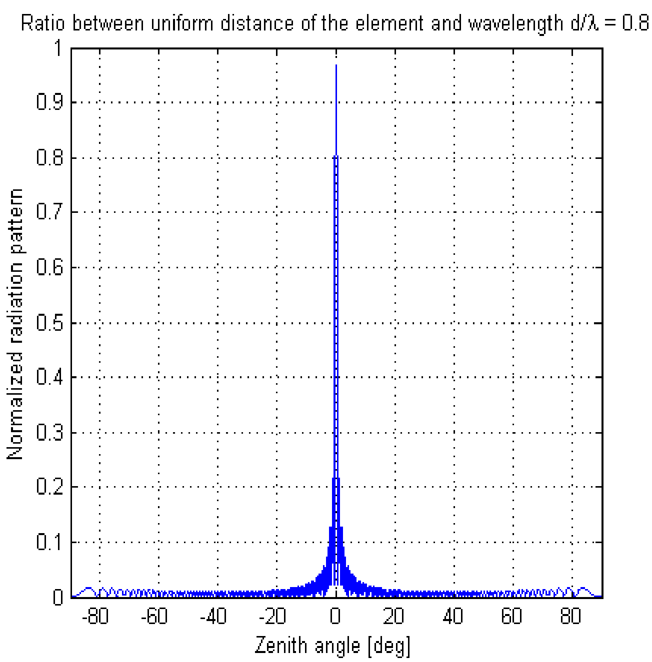


Fig. 5 Radiation pattern when ratio between distance and wavelength d/ λ =0.8

When we switch to the ratio of d/λ from 0.3 to 0.8 we get our desire beam pattern by suppressing almost all of the side

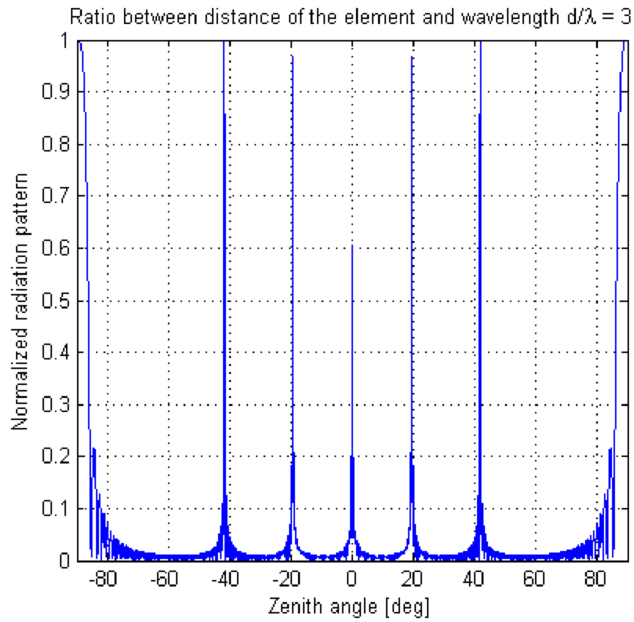


Fig. 6 Radiation pattern when ratio between distance and wavelength d/ λ =3

lobes. For the higher efficiency from the phased array antenna for SPS system we need narrower beam width with reducing all of the side lobes. From Fig.5 we can say that normalized radiation pattern of main beam at the center is thin enough keeping the side lobes almost at zero level which is the best possible result for 100 elements phased array antennas for SPS system from the simulation result.

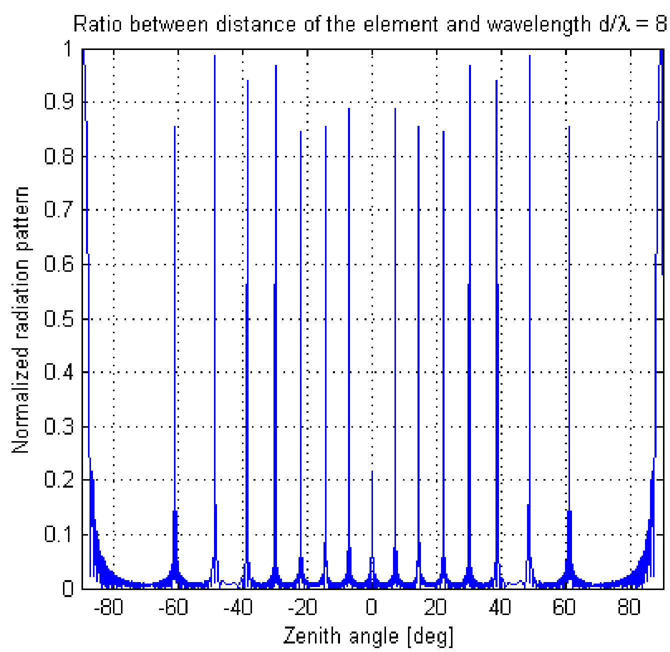


Fig. 7 Radiation pattern when ratio between distance and wavelength d/ λ =8

Moreover, we have simulated the radiation pattern for the phased array of 100 elements for d/λ =3 and d/λ =8 shown in the Fig.6 and Fig.7 respectively where we can see that there are several side lobes having normalized radiation patterns much bigger than the main lobe. When we switch into the ratio of d/λ =8 shown in figure 7 we find normalized radiation patterns of main lobe is in the range of 0.21 where side lobes are between the ranges of 0.85 to 1 which is not acceptable. According to the above simulation result we can conclude that for the 100 element uniform linear phased array antenna, d/λ =0.8 is the best possible radiation pattern for SPS system.

V. CONCLUSION

In search for the alternative source of rapidly dwindling fossil fuel resources and for the replacement of the hazardous nuclear power plant Space based Solar Power Satellite system is a potential solution which will provide a consistent, stable and renewable source of energy for a long time once the initial investment is made. Phased array antenna is the significant part of the SSPS system for the highly efficient microwave power transfer. From our observation it is asserted that for the 100 elements phased array antenna if we can keep the ratio between uniform distance of the phased array element and wavelength of the signal in the range of 0.8 then we will get the desire radiation pattern of the main beam width keeping the side lobes almost in the zero level. However, if we can overcome the economical constraint to launch SSPS system as well as practical implementation of efficient phased array antenna SSPS system will minimize highly demandable power crisis for the developing country as well as for the world. Our obvious area of future improvement will be to investigate the highly radiation pattern of phased array increasing the total elements which may include further analysis of time delay elements.

REFERENCES

[1] Golap Kanti Dey, Kazi Tanvir Ahmmed; "Multi-Junction Solar Cells and Microwave Power Transmission Technologies for Solar Power Satellite", 3rd IEEE International Conference on Informatics, Electronics & Vision, pp-1-6, 2014.

[2]R. C. Hansen, "Phased Array Antennas" Second Edition, Copyright-2009 by John Wiley & Sons, Inc., 2009.

[3] Hsi-Tseng Chou , Chien-Te Yu "Design of phased array antennas with beam switching capability in the near-field focus applications" IET Microwave Antennas Propagation., Vol. 9, Issue. 11, pp. 1120–1127, 2015.

[4] Takenori Yasuzumi, *et al.* "A Study on Phased Array Antenna Using MSA Composed by 3-patches", Loughborough Antennas & Propagation Conference UK, pp. 281-284. 2009.

[5] Tuan Thanh Ta, *et al.* "A 3-D Radiation Pattern Measurement Method for a 60-GHz-Band WPAN Phased Array Antenna" , Proceedings of APMC 2012, Kaohsiung, Taiwan, pp. 139-141 , 2012.

[6] Takaki Ishikawa, Yuta Kubo, Junki Yoshino and Naoki Shinohara "Study of Beam Forming for Microwave Power Transmission toward Solar Power Satellite with Advanced Phased Array System in Kyoto University", 2013 IEEE Antennas and Propagation Society International Symposium (APSURSI), 2225 – 2226, 2013.

[7] Gregory M. Turner, Christos Christodoulou, "FDTD Analysis of Phased Array Antennas", IEEE Transactions on Antennas and Propagation, Vol. 47, No. 4, 1999.

[8] Hiroyuki Kayano, *et al.* "Low-Profile High-Sensitivity Sub-array Module with HTS Filters for an Active Phased Array Antenna", IEEE Radar Conference, pp. 118-121, May 2014.

[9] N. C. Karmakar , M. E. Bialkowski, "Electronically Steerable Array Antennas for Mobile Satellite Communications-a review," Proceedings of the IEEE International Conference on Phased Array Systems and Technology, Dana Point, CA, USA, pp. 81–84, 2000.

[10] D. Ehyaie, Novel approaches to the design of phased array antennas (Ph.D. thesis), University of Michigan, Ann Arbor, Mich, USA, 2011.

[11] Matthew G. Bray, *et al.* "Optimization of Thinned Aperiodic Linear Phased Arrays Using Genetic Algorithms to Reduce Grating Lobes During Scanning" , IEEE Transactions on Antennas and Propagation, Vol. 50, No. 12, pp. 1732-1742, 2002.

[12] Matthew G. Bray, *et al.* "Thinned aperiodic linear phased array optimization for reduced grating lobes during scanning with input impedance bounds", IEEE Antennas and Propagation Society International Symposium, Boston, USA, Vol. 3, pp. 688-691,2001.

Cognitive Radio Enabled VANET for Multi-agent Based Intelligent Traffic Management System

S. M. Nadim Uddin[1], Nafees Mansoor[2], Sazzad Hossain[3]
[1] Department of Electronics and Telecommunication Engineering, University of Liberal Arts Bangladesh
[2,3] Department of Computer Science and Engineering, University of Liberal Arts Bangladesh
E-mail: [1]sm.uddin.ete@ulab.edu.bd, [2]nafees.mansoor@ulab.edu.bd, [3]sazzad.hossain@ulab.edu.bd

***Abstract*—With the mounting interest on cognitive radio (CR) technology in wireless communication systems, it is anticipated that CR-enabled vehicular networks will play a vigorous role in the enrichment of communication efficiency in vehicular network. This paper presents a Cognitive Radio enabled VANET for multi-agent based intelligent traffic management system. A skeleton for intelligent learning and decision mechanism for Central Traffic Management is also proposed and discussed in the model. The proposed model has two distinct information exchange system layouts. One is dynamic (vehicle to vehicle) and another is semi-dynamic (vehicle to Road-Side-Unit). For the vehicle-2-vehicle communication, the proposed model assumes that vehicles can communicate with each other using available wireless resources with the help cognitive radio mechanism. This paper also introduces a cluster formation scheme for better accuracy in data transmission among vehicles. The dynamic module of the proposed model is later simulated and validated for some important performance communication metrics.**

Keywords—Cognitive Radio; VANET; Traffic Management System; CR-VANET; MANET

I. Introduction

With the advancement of microelectronics and communication modules, a rapid surge of interest is observed in the research community for Mobile Ad-hoc Network (MANET) and Vehicular Ad-hoc Network (VANET). MANET, a special type of ad-hoc network, is self-organized and is fully operable without the assistance from any fixed infrastructural support or central administration. The mobile nodes are inter-connected by the wireless links and communication is held directly between nodes or through intermediate nodes [1]. Each node in MANET continuously maintains the updated information that is required to properly route the traffic [2]. The Internet Engineering Task Force (IETF) has developed two standard track routing protocols for MANET, namely proactive and reactive MANET protocols [3].

On the other hand, VANET is a distinctive class of MANET where moving vehicles act either as nodes for direct communication or as routers to provide intermediate connectivity. These vehicular nodes can communicate with other vehicles to establish Vehicle-to-Vehicle communication system. These nodes can also communicate with the access point (AP) to establish an Infrastructure-to-Vehicle communication system. A VANET network consists of four major components, namely Vehicles (nodes or mobile hosts), On-Board Unit, Road-Side Unit and Central management system [5].

Thus, VANET follows and applies the same principle of MANET in a highly dynamic environment of surface transportation. However, due to the high mobility of the nodes, VANET needs to consider dynamic information exchange and unreliable channel conditions. These considerations are absent in MANET. As a result, deployment of MANET protocols in VANET show poor performance [6]. Since the vehicles move along the road, movements of the vehicles are predictable in VANET, where network density changes over time and location [4].

According to the IEEE 802.11 independent basic service set (IBSS), no access point is required to communicate in distributed peer-to-peer manner and IBSS operation can occur if two nodes are within the radio range of each other. [5].

However, the radio spectrum scarcity has become more vigorous concern with the amplifying demand in wireless applications. To combat the growing demand of radio spectrum, proper utilization of the radio spectrum is essential. Cognitive radio exercises dynamic spectrum allocation technique to utilize radio spectrum efficiently and reduces the bottleneck on frequency bands. With the recent advances in cognitive radio systems [23-24], cognitive radio enabled vehicular users in VANETs would be able to sense and hop from one frequency to another (or one network to another) in the entire spectrum range based on their needs and operating environment with the help of cognitive radios. While several studies exist in literature on applying CR to wireless mesh networks, ad hoc networks, and cellular net- works, the research on applying CR to VANETs is still in its early stage. The research solutions proposed for general-purpose CR networks cannot be directly applied to CR-VANETs as the unique features of vehicular environment, such as the role of mobility, and the cooperation opportunities need to be taken into account while designing the spectrum management functions for CR-VANETs [25-26].

In this paper, Cognitive Radio Enabled VANET for multi-agent based intelligent traffic management system is proposed. The proposed system ensures data exchange between high-speed vehicles and between the vehicles and the roadside infrastructure in the licensed ITS band (5.85-5.925 GHz). In this model, vehicles can communicate with each other using available wireless resources with the help cognitive radio mechanism. Without the loss of generality, the proposed model is simulated and validated for some important performance communication metrics. In this paper, a cluster formation mechanism is considered for better accuracy in data transmission among vehicles. A skeleton for intelligent learning and decision mechanism for a Central Traffic Management is also proposed and discussed in the scope of the model.

The paper is organized as follows. In section II, a brief analysis on different lately developed protocols for VANET Communications is discussed. The proposed network model and architecture for "Cognitive Radio Enabled VANET for Multi-agent Based Intelligent Traffic System" are described in section III. In section IV, the simulation results of the proposed model are presented. Conclusion and future works have been discussed in section V.

II. RELATED WORKS

Different protocols have been proposed for better performance in wireless networks for vehicular communications. OLSR sends two types of messages namely hello message and Topological control message [7]. This protocol gives better performance among the proactive routing protocols [8]. DSDV uses the shortest path to find the route to the destination and guarantees loop free nodes reduces count to infinity problem and also reduces control message overhead. This protocol is suitable only for smaller number of nodes [9].

Ad-hoc on-demand distance vector (AODV) protocol discovers routes only on demand i.e. it establishes a route only when any node needs to send a message to the destination. It offers low network overhead by avoiding the flooding of messages periodically in the network. It requires less memory size and the routing tables only contain the recent active nodes. AODV is flexible to highly dynamic and large-scale network [10]. Ad-hoc On-demand Multipath Distance Vector Routing (AOMDV) protocol maintains multiple loop free path with minimum overhead. It is suitable for high mobility nodes [11]. Dynamic Source Routing (DSR) protocol mainly consists of two mechanisms, namely route discovery and route maintenance and uses a unique id request in the route request packet [13]. Temporally Ordered Routing Algorithm (TORA) protocol uses multi hop routes. This protocol is based on the link reversal routing algorithm which uses directed acyclic graph to identify the flow of packets . TORA's performance is better than DSR in highly dynamic ad-hoc environment [13].

Zone Routing Protocol (ZRP) is the first hybrid routing protocol which divides the network into overlapping zones. It uses the proactive routing scheme inside the zone and reactive routing scheme outside the zone [14]. Core Extraction Distributed Ad-hoc Routing (CEDAR) is a protocol with integrated QoS support [15]. Distributed Dynamic routing algorithm Protocol (DDR) is a tree based routing protocol that does not require the root node support for data transfer. Greedy Perimeter Stateless Routing (GPSR) selects node closer to the destination using beacon [17]. Greedy Perimeter Coordinator Routing (GPCR*)* is a position-based overlay routing protocol that uses greedy algorithms to forward packet based on a pre-selected path .It has been designed to meet the challenges of city scenarios. No Global Information System required for GPCR [18]. Connectivity- Aware Routing (CAR) is well suited for city and highway scenarios. It uses AODV for path discovery and PGB for data dissemination. It also uses guard concept for path maintenance. It ensures the shortest connected path and no digital map is required for CAR. It has higher packet delivery ration than GPSR [18]. GSR (Geographic Source Routing) is designed for city environment, uses greedy forwarding approach along pre-selected path using Djkshtra's shortest path algorithm. It combines the features of topological information and position based routing [19]. Recent cognitive radio based wireless mesh network related work either use the channel selection within a network [10] or transmit power or rate adaptation within a given network [11]. These types of solutions are not directly applicable in heterogeneous wireless environment since different networks have different characteristics. All of the aforementioned research works are related to protocol designing for VANET & protocol proposals and spectrum sensing for CR.

However, there is no cluster-based ad-hoc routing protocol-integrated intelligent surface traffic management system has been proposed.

III. PROPOSED MODEL

A. Network Model

In the proposed model, it is assumed that there are two distinct scenarios in terms of nodes' density namely highly dense network and light dense network. A mobile vehicle is noted as node where mobile vehicles can be cars, buses etc. in the network. A sender node is defined as a particular node from where the data packets are coming and destination node is defined as the sink or the desired recipient of the data packets. A grid is defined as a geographical area with at least a Roadside Unit (RSU) and a predefined number of nodes with On-Board Units (OBU) installed into them. Here node density is the number of nodes in a grid and network is considered as the accumulation of all grids. A RSU is a hardwire mounted on top of a pole or tower in a grid, which is capable of maintaining simultaneous wireless duplex connections with nodes and Central Traffic Management Unit (CTMU). RSUs build a neural network with each other and the CMTU.

In a grid, there are two types of networks, which is labeled as Dynamic networks and Semi-dynamic networks. Dynamic network is created among nodes where nodes can move along a predefined manner and semi-dynamic network is created between a node and a RSU.

In a dynamic network, the nodes can communicate with each other using the principle of cognitive radio with the help of OBUs, which follow the IEEE 802.11p protocols. OBUs are connected with each other via wires or wireless networks. CMTU is the central management system, which periodically updates traffic conditions and sends relevant data to RSUs for proper route maintenance and nodal information to be displayed in Dynamic Info Board (DIB). A DIB is an info board from where nodal position and approximate speed and arrival time can be seen. It is mounted on the same pole of a RSU.

Data packet, in the model, follows a basic structure consisting of sender node id, destination id, hop count, timer, sequence number etc. The routing path for source to destination is determined by the greedy forwarding algorithm which means node closest to the destination node among all neighboring nodes in the transmission range of the sender node is selected as the next relaying node or the next hop. It is assumed that MAC protocol is TDMA, in which time is slotted and synchronized and to ensure proper sharing of the wireless resources, an appropriate scheduling algorithm is selected. For simplicity, rerouting mechanism (in case of link failure), packet collision probability (two nodes trying to send data packets to one node at the same time), nodes' mobility (speed) etc. are not considered in the scope of this paper.

B. Architecture:

The proposed model is an intelligent traffic management system mainly focused on city scenarios though it can be implemented in highways as well as rural areas.

In an arbitrary geographical location, it is assumed that there are some grids with some fixed RSUs. When a node enters a grid, it is automatically connected to the wireless network of that grid through its OBU. OBU then sends data packets to its nearby nodes or RSU. In a wireless ad-hoc network, a node can only send data packets if there is another node or the desired node within the transmission range of the sender node.
If the RSU is not in the range of the sender node, the data packets can be sent from the sender node to the destination node or RSU via intermediate nodes, which are labeled as relaying nodes.

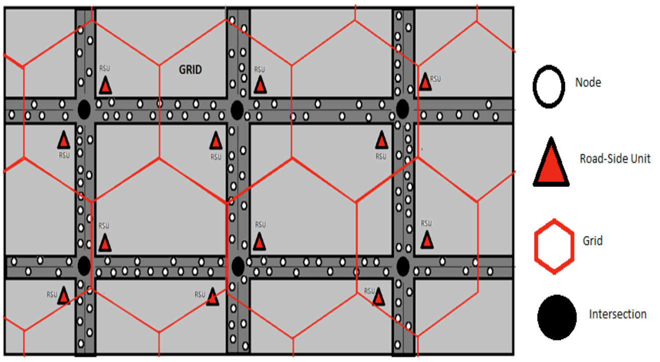


Fig 1: The proposed traffic management system

A RSU picks up the data packets sent to it and extract the necessary information for traffic conditions in that grid and sends the data to CTMU for further processing. CMTU then processes the received data and sends the updated data to the RSUs. RSUs then display information of selected vehicles (local transports, ambulance etc) such as positions and approximate arrival time in DIB. An overview of advanced traffic system is shown in Fig 1.

Suppose a node “A”, which is a local transport, enters a grid namely “XYZ”. In XYZ, there are several nodes “B”, “C”, “D”, “E” etc. “A” is then connected to a network label as “GHJ123” in which all other nodes are connected as well. In that grid, there is a RSU, labeled as “RSU1”, at an intersection. “A” needs to send data packets to RSU1, but it is not in the range of “A”. So, the data packets from A can be sent to RSU1 via B, C and D. RSU1 receives the data packets and send it to the CTMU. CMTU processes data and sends the updated data to the RSU1, RSU2, RSU3 etc. A basic network with links among nodes and RSU is shown in Fig 2.

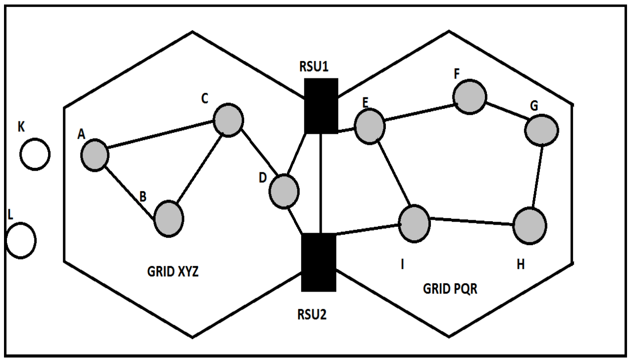


Fig 2: Node view

According the mechanism discussed in the Network Model subsection, clusters form among the nodes to achieve better accuracy in data packet exchange. In a cluster, all links among the nodes have link-weight, which depend on the possession of available channels in cognitive networks. The node with greater link-weight is selected as the "Cluster head" and other nodes within the transmission range of the CM become members of that particular cluster. Cluster-head becomes responsible of transmitting data packets to the destination or to the next cluster through "Edge Members". Edge members are responsible for maintaining the links among clusters. To avoid data packets loss, a "Secondary Cluster head" is selected in case of unavailability of Cluster head. An example of cluster formation among nodes is illustrated in Fig 3.

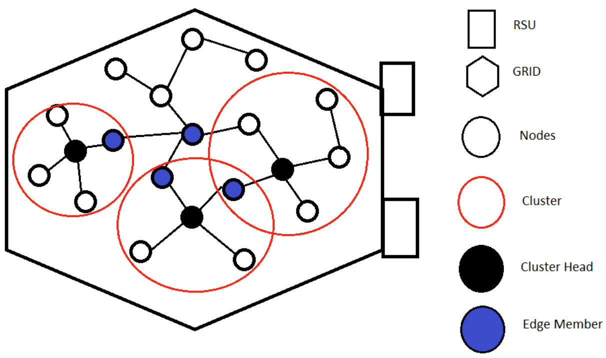


Fig 3: Proposed clustering scheme

On-Board Unit is a hardware implemented on the vehicles, which creates and maintains links with other nodes or RSU. It operates according to IEEE 802.11 protocols and it utilizes the available wireless resources such as bandwidth, channels etc. with the help of cognitive radio network mechanism. Though the OBU will continuously consume energy as it is active as long as the vehicle is active, it is assumed that the node has sufficient power supply for it. The main task of the OBU is to sense spectrums and maximize the possibility to be online by using or sharing available bandwidth so that it can broadcast its data packets to the destination node.

A DIB is a board, which displays some basic real time information about the partial grid traffic and local transports. It is connected to the RSU on which it is mounted. Whenever the RSU is updated from the CMTU, updated information about some specific vehicles shows up on the DIB. It shows approximate location, time of arrival, accidents etc.

CMTU is the main server or the accumulation of several main servers, which controls the whole traffic system. It is connected with the RSUs via wire or wireless connections and builds a neural network. CMTU has an intelligent learning and decision-making algorithm, which extracts necessary data from the data, received from the RSUs and updates its database for future decision making. CMTU then updates the RSUs and the specific updated data for specific route is shown in the DIB of specific RSU. CMTU allocates time intervals in intersections based comparison in previous data and the updated data in its database.

C. Cluster formation

With the mounting interest on cognitive radio (CR) technology in wireless communication systems, it is anticipated that CR-enabled vehicular networks may improve the vehicular communication efficiency. Thus, considering the possibilities of cognitive radio technology, the dynamic module of the proposed network model is designed to be functional on cognitive radio environment. As discussed earlier in the network model section, the dynamic vehicular network in each grid is divided into some sub-groups or clusters. Clustering concept is introduced in the dynamic module as cluster-based ad-hoc network aims to achieve better accuracy in data packet exchange.

The proposed clustering mechanism is inspired from author's previous clustering scheme for cognitive radio ad-hoc network [20-22]. In the existing clustering mechanism, cluster-head selection is based upon a weight, where to calculate the weight number of common channels and number of neighboring nodes is taken into consideration. However, in this paper another parameter has been taken into account called node's speed along with the previous two to calculate weight for each node.

Once the weight calculation for each node is completed, the node with higher weight is selected as the "Cluster head" and other nodes within the transmission range of the Cluster-head (CH) become cluster members of that particular cluster. Cluster-head becomes responsible of transmitting data packets to the destination or to the next cluster through "Edge Members". Edge members are responsible for maintaining the links among clusters. To avoid data packets loss, a "Secondary Cluster head" is selected in case of unavailability of Cluster head. An example of cluster formation among nodes is illustrated in Fig 3.

In the proposed clustering scheme, CH defines and upholds operating channels for the cluster. To find the existence of any other clusters in the neighborhood, CMs check their neighbor list for other cluster heads. CM becomes the Edge Member (EM) and connects two clusters once it finds other CH in the neighbor list. In the proposed clustering scheme, cluster consists of one CH, one SCH and CMs. All cluster members are 1-hop apart from the CH. EM connects two neighboring clusters, where there can be maximum two intermediate EMs between two CHs. Using local common channels, intra-cluster communications are performed.

IV. SIMULATION RESULTS

A. Simulation Environment

To evaluate the clustering performance of the dynamic module (vehicle-2-vehicle) of the proposed model, simulation is conducted. Though several network simulators are available, whose output depicts as close as possible to real time implementation, to simulate and analyze performances of the proposed model, discrete-event simulator NS2 has been used and the performance analysis are conducted using PERL scripts.

Four parameters, namely throughput, energy consumption, delay and overhead are considered as the performance metrics to evaluate the performance of the network. Moreover, two distinct dynamic scenarios in terms of nodes' density are considered for the comparative study in the simulation environment. Thus, in one scenario the radio transmission range of a node is considered to be 100 meters and in the other scenario, the transmission range is set to 500 meters. For both scenarios, number of nodes is considered to be 100 where distributed sources and sinks are altered randomly.

A simulation area of 4000 m^2 is considered for the simulation purpose. The Two-Ray Ground model is used as the propagation model and Drop-tail method is used for the queuing purpose. MAC/802.11p is considered as the MAC type and maximum packet queuing delay is considered as 50ms. In the simulation, data traffic is generated with Constant Bit Rate (CBR) with packet size is set to 512 bytes. Varied packet rate ranging from 100 packets/sec to 800 packets/sec is considered to evaluate the performance of the network for different traffic load. Initial energy for all the nodes is considered to be 10 Joules. Maximum speed for the nodes is considered to be 10 m/s. The simulations are run for 150s each and the results are calculated as mean of several observations in light dense dynamic environment and heavily dense dynamic environment.

B. Performance Evaluation

This section of the paper discusses the simulation results of the proposed model in terms of throughput, energy consumption, delay and overhead.

1. Performance based on Throughput

In this paper, throughput is defined as the number of successfully received data packets at the destination node in a unit time and it is represented in Kbps. Maximum throughput is preferable while designing the routing protocol for the proposed dynamic network. In Fig. 4, the horizontal axis indicates the traffic load and the vertical axis indicates the throughput.

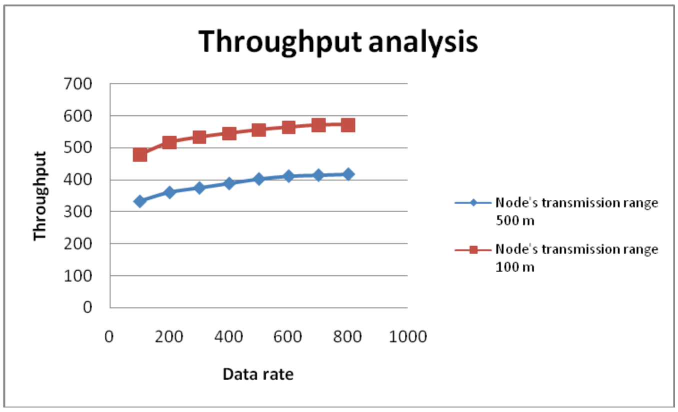


Fig 4: Performance Evaluation in Terms of Throughput

Fig. 4 shows that with an increasing traffic load, throughput for both scenarios increases. However, throughput is higher in the network with the radio transmission range of 500 meters, than the radio transmission range of 100 meters for all different data flow rates. This is because; a network with long ranged transmission finds lesser number of hops to transmit packets to the destination. Thus, with decreasing number of hops, number of links throughout the network, probability of link failure and rate of packet retransmission reduce significantly. Therefore, with higher transmission range, throughput of the network increases for all different traffic loads.

2. Performance based on Energy consumption

In this paper, energy consumption is defined as the cumulative sum of consumed energy by all the nodes of the network during the entire transmission period, where consumed energy of a node is calculated by subtracting remaining energy from initial energy. The unit for energy consumption has been considered as Joule. Minimum energy consumption is desirable while designing the routing protocol for the proposed dynamic network. In the simulation environment, consumed energy of a node mainly depends on the functional period of a node to transmit the data packets to its next hop. Moreover, network energy consumption also depends on the number of relying nodes while transmitting packets. In Fig. 5, the horizontal axis indicates the traffic load and the vertical axis indicates the consumed energy.

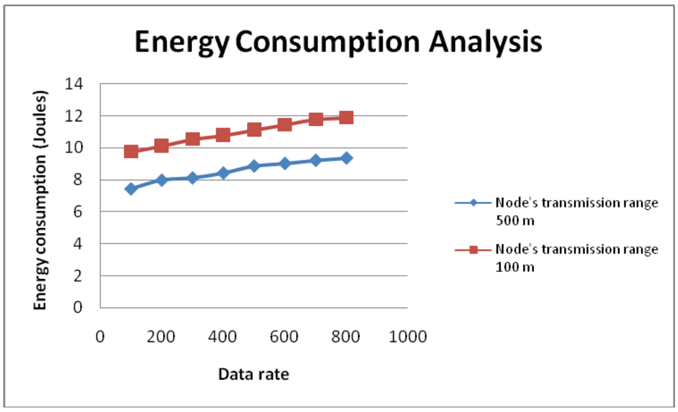


Fig. 5: Performance Evaluation in Terms of Consumed Energy

From the figure, it is seen that the energy consumption increases with increasing data flow rate in both scenarios. The reason is that when traffic load is increased, more data packets need to be transmitted over the network from the source node to the destination node, where the intermediate nodes are required to be at the active state for longer period of time. As consumed energy of a node is considered to be dependent on the functional period, therefore with increasing traffic load, energy consumption is also increased in both networks.

Moreover, it can also be seen from Fig. 5 that the energy consumption is lesser in a network with radio transmission range of 500 meters compared to the network with radio transmission range of 100 meters for all different traffic loads. This is because, when radio transmission range in a network is longer, number of hops from the source to the destination tends to be lesser than that of a network of shorter transmission ranged radios. Thus, a network with long ranged transmission results lesser number of intermediate nodes between a source and the destination than the network with short ranged transmission. Therefore, energy consumption remains lesser in a network with longer transmission ranged radios as lesser number of intermediate nodes is engaged to forward the data packet from the source node to the destination.

3. Performance based on Packet Transmission Delay

In this paper, the packet transmission delay is defined as the average time required for transferring data packets from the source node to the destination node. Minimum packet transmission delay is preferable while designing the routing protocol for the proposed dynamic network. In Fig 6, the horizontal axis indicates the traffic load and the vertical axis indicates the delay, where it is seen that the delay varies with varying data flow rate in both scenarios.

From the figure it is observed that the packet transmission delay increases with increasing data flow rate in both scenarios. The reason is that when traffic load is increased, more data packets need to be transmitted from the source to the destination node. As a result, the intermediate nodes are required to process more packets which eventually increases individual data processing sessions among the nodes. Thus, when higher number of packets propagates, source node and the intermediate nodes need longer time to forward the packets to the next hop, which increases the cumulative packet transmission delay.

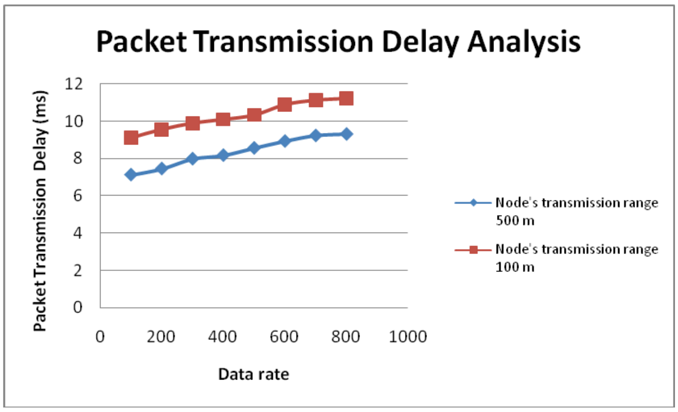


Fig. 6: Performance Evaluation in Terms of Packet Transmission Delay

Moreover, it can be also observed from Fig. 6 that the packet transmission delay is lesser in a network with radio transmission range of 500 meters than in the network with radio transmission range of 100 meters for all different traffic loads. This is because, when radio transmission range in a network is longer, number of hops from the source to the destination node is reduced than that of a network with shorter transmission ranged radios. Thus, lesser number of intermediate nodes is engaged to forward the data in the network, which results lesser data processing sessions. Therefore, a network with long transmission ranged radios results lesser packet transmission delay than that of a network with short transmission ranged radios.

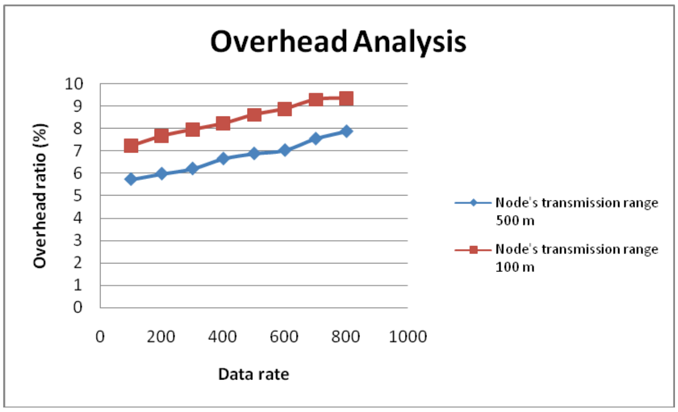


Fig. 7: Performance Evaluation in Terms of Overhead Ratio

4. Performance based on Overhead Ratio

The network overhead is considered as the sum of transmitted control packets during the transmission sessions. Thus, network overhead ratio is defined as the total transmitted control packets over total received data packet at the destination node. In Fig. 7, the network overhead ratio is presented in terms of percentage, where network overhead ratio is desirable to be minimum while designing the routing protocol for the proposed dynamic network. In the figure, the horizontal axis indicates the traffic load and the vertical axis indicates the network overhead ratio.

From the figure, it is seen that the overhead ratio increases with increasing data flow rate in both scenarios. The reason is that when traffic load is increased, more data packets need to be transmitted over the network from the source node to the destination node. In a wireless ad-hoc network, each node has a defined data packet queuing delay. Increase in data rate means nodes need to process more data packets individually which eventually results in reduction of a node's efficiency for packet forwarding. As a result, with the increase of traffic load, retransmission of data packets from source node to destination node increases and therefore, acknowledgement control messages for retransmission to source node and from destination node increase, network overhead increases. Thus, with the increase in traffic load, the network overhead ratio increases.

Moreover, it is also observed from Fig. 7 that the overhead ratio is lesser in a network with radio transmission range of 500 meters compared to the network with radio transmission range of 100 meters for all different traffic loads. That is because, nodes in a network with shorter radio transmission range will need more number of relaying nodes for transmitting data packets to the destination node from the source node than that of with longer radio transmission range. As the number of the nodes is comparatively more, more nodes will have to process increased traffic load and as a result, retransmission of data packets due to exceeding node's capacity to process data packets will increase which eventually increases number of transmitted control packet. Thus, a network with radio transmission range of 100 meter will have more network overhead ratio than that of a network with radio transmission range of 500 meters.

V. CONCLUSION AND FUTURE WORKS

This paper presents a Cognitive Radio enabled VANET for multi-agent based intelligent traffic management system. The proposed model has two distinct information exchange system layouts, namely dynamic module (vehicle to vehicle communication using cognitive radio) and semi-dynamic module (vehicle to Road-Side-Unit). For the dynamic module, a cluster formation scheme is introduced and later simulation is conducted to validate the performance. Our next research steps are to develop the prototype of the model and to develop an intelligent algorithm that can give optimal decision to manage the traffic.

REFERENCES

[1] Sesay, S., Yang, Z., & He, J. (2004). A survey on mobile ad hoc wireless network. Information Technology Journal, 3(2), 168-175.

[2] Timcenko, V., Stojanovic, M., & Rakas, S. B. (2009, August). MANET routing protocols vs. mobility models: performance analysis and comparison. In Proceedings of the 9th WSEAS international conference on Applied informatics and communications (pp. 271-276).

[3] Retana, A. Mobile Ad-hoc Networks, charter-ietf-manet-03 (1997)

[4] Anwer, M. S., & Guy, C. (2014). A survey of VANET technologies. *J. Emerg. Trends Comput. Inf. Sci*, *5*, 661-671

[5] Nadeem, A., & Howarth, M. P. (2013). A survey of MANET intrusion detection & prevention approaches for network layer attacks.Communications Surveys & Tutorials, IEEE, 15(4), 2027-2045.

[6] Li, F., & Wang, Y. (2007). Routing in vehicular ad hoc networks: A survey.*Vehicular Technology Magazine, IEEE*, *2*(2), 12-22

[7] Bilal, S.M., C.J. Bernardos, and C. Guerrero, Position Based Routing in Vehicular Networks: A Survey. Journal of Network and Computer Applications, 2012.

[8] Philippe Jaquet, Paul Muhlethaler and Amir Qayyum,―Optimized Link State Routing Protocol‖, IETF

[9] Jamal toutouh,Jose Garcia-Nieto and Enrique Alba,Intelligent OLSR routing Protocol Optimization for VANETs,IEEE transactions on vehicular technology,vol.61,No.4,May 2012

[10] Bai, Y., S.B. Ji, and Y. Xue, A Comparative Study of AODV and DSDV Routing Protocol of the Vehicle Ad Hoc Network Based on NS2.

[11] C. E. Perkins and E. M. Royer, ―Ad Hoc On-demand Distance Vector Routing‖, In Proceedings of the 2nd IEEE Workshop on Mobile Computing Systems and Applications, New Orleans, LA, 1999, 90-100.

[12] D. B. Johnson, D. A. Maltz, Yih-Chun Hu and Jorje tagJetcheva, ―The Dynamic Source Routing for Mobile Ad Hoc Wireless Networks‖, IETF Internet draft, Nov.2001.

[13] Pirzada, A.A., C. McDonald and A. Datta, 2006.Performance comparison of trust-based reactive routing protocols. Mobile Computing, IEEE Transactions on, 5(6): 695-710.

[14] Kashif Naseer Qureshi and Abdul Hanan Abdullah , Study of Efficient Topology Based Routing Protocols for Vehicular Ad-Hoc Network Technology, World Applied Sciences Journal 23 (5): 656-663, 2013

[15] Karp, B. and Kung, H. T (2000), ―GPSR: greedy perimeter stateless routing for wireless networks. In Mobile Computing and Networking, pages 243-254, 2000.

[16] Naumov, V., Baumann, R., Gross, T. (2006), ―An evaluation of Inter-Vehicle Ad Hoc Networks Based on Realistic Vehicular Traces, Proc. ACM MobiHoc'06 Conf., May, 2006.

[17] Lochert, C., Mauve, M., F¨ussler, H., and Hartenstein, H., ―Geographic routing in city scenarios,‖ SIGMOBILE Mob. Comput. Commun. Rev., vol. 9, no. 1, pp. 69–72, 2005.

[18] Naumov, V., Gross, T.R. (2007), "Connectivity-Aware Routing (CAR) in Vehicular Ad-hoc Networks," INFOCOM 2007. 26th IEEE International Conference on Computer Communications. IEEE , vol., no., pp.1919-1927, 6-12 May, 2007.

[19] M. Jerbi, S. M. Senouci, R. Meraihi, Y. Ghamri-Doudane, ―An improved vehicular ad hoc routing protocol for city environments, In Communications, 2007, pp. 3972-3979.

[20] Mansoor, N., Islam, A. K. M., Zareei, M., Baharun, S., & Komaki, S. (2013, December). Spectrum aware cluster-based architecture for cognitive radio ad-hoc networks. In Advances in Electrical Engineering (ICAEE), 2013 International Conference on (pp. 181-185). IEEE.

[21] Mansoor, N., Islam, A. M., Zareei, M., Baharun, S., Wakabayashi, T., & Komaki, S. (2015). Cognitive radio ad-hoc network architectures: a survey. Wireless Personal Communications, 81(3), 1117-1142.

[22] Mansoor, N., Islam, A. K. M. M., Zareei, M., Baharun, S., & Komaki, S. (2014). Cluster modelling for cognitive radio Ad-hoc networks using graph theory. In 2014 International Conference on Applied Mathematics, Modelling and Simulation (ICAMMS) (pp. 1-8).

[23] G. Marfia, M. Roccetti, A. Amoroso, M. Gerla, G. Pau, and J. H. Lim, "Cognitive Cars: Constructing a Cognitive Playground for VANET Research Testbeds," in 4th International Conference on Cognitive Radio and Advanced Spectrum Management (CogART 2011, Barcelona, Spain, Oct. 2011

[24] D. B. Rawat, B. B. Bista, and G. Yan, "CoRVANETs: Game Theoretic Approach for Channel and Rate Selection in Cognitive Radio VANETs," in Proceedings of the Seventh International Conference on Broadband and Wireless Computing, Communication and Applications (BWCCA-2012), Nov 2012

[25] Yan, M. Li, Z. Yang, W. Lou, and H. Zhai, "Throughput Analysis of Cooperative Mobile Content Distribution in Vehicular Ad Hoc Networks using Symbol Level Network Coding," IEEE Journal on Selected Areas in Communications (JSAC), special Issue on Cooperative Networking: Challenges and Applications, 2012.

[26] P. Shankar, T. Nadeem, J. Rosca, and L. Iftode, "CARS: Context-Aware Rate Selection for vehicular networks," in Proceedings of the IEEE International Conference on Network Protocols, 2008. ICNP 2008., November 2008, pp 1-12

Network Formation and Data Centric Routing in Wireless Sensor Networks

Asim Zeb[1], Javed Bangash[2]
[1]Qurtaba University, Peshawar, Pakistan
[2]Abasyn University, Peshawar, Pakistan

A.K.M Muzahidul Islam, Sabariah Baharun, Atiqur Rahman
Malaysia Japan International Institute of Technology, Universiti Teknologi Malaysia, Malaysia.

Yoshiaki Katayama
Graduate School of Engineering, Nagoya Institute of Technology, Nagoya, Japan.

Abstract--- **Research on Wireless Sensor Network (WSN) has been increased tremendously in recent years. This paper investigates problems associated with network maintenance and data routing in dynamic cluster-based wireless sensor network (DCWSN) architecture. This article addresses multiple nodes joining to DCWSN under consideration of single channeled network and Data Centric Routing. Multi nodes joining with complete 1-hop neighbor information (MNJ-CN) is developed to join multi nodes. Moreover, Data Centric Routing algorithm is also used to adapt the routing strategy. The routing strategies are of two considerations such as, reliability and delay. To evaluate efficiency of the algorithms, simulation experiment is conducted. Simulation results show that both the proposed algorithms, namely MNJ-CN and Data Centric Routing perform more efficient than existing algorithms.**

Keywords-- Dynamic, Wireless Sensor Network (WSN), network maintenance, Data routing, Reliability, delay.

I. INTRODUCTION

Wireless sensor network is composed of sensor nodes. Besides environment monitoring, node (sensor) also passes messages to destination node(s) using communication channel(s). Sensor nodes have limited resources in terms of power transmission, communication range and bandwidth. Different applications of wireless sensor networks are robot control, smart home, offices, and manufacturing [1, 2].

Network clustering is a widely practiced solution to achieve efficient network performance. In a cluster-based network, cluster head (CH) is a local coordinator that communicates with member nodes (MN) and other clusters. MNs are the leaf nodes that usually operate around cluster head. MNs gather data from environment and send to the cluster head. Two key activities performed in cluster-based WSNs namely, cluster formation and cluster maintenance. Cluster formation handles the development of clusters. On the other hand, cluster maintenance is an optional and important activity in cluster-based network. Cluster maintenance makes the network behavior dynamic by ensuring stability in the network in case of node joining and leaving. The absence of cluster maintenance may rebuild the network from scratch when a new node joins or existing node leave the network [4]. Therefore the complexity of cluster maintenance may equal to cluster formation that highly degrades network performance. Consequently, the network maintenance is crucial and is an important procedure to maintain the integrity of the network [3]. Moreover, cluster maintenance proactively handle the topological manipulation and reconfiguration without any data lose [5- 6].

Cluster maintenance presumes when new nodes are deployed in the network or existing node leave the network. The objective of this paper is to build an efficient cluster maintenance scheme for node joining and data routing scheme. In order to join multi nodes the network, multi nodes joining with complete 1-hop neighbor information (MNJ-CN) is also proposed. MNJ-CN is the improvement of single node joining algorithm [12-13]. Minimizing number of messages is the advantage of MNJ-CN is another advantage of multi nodes joining. MNJ-CN algorithm also requires lesser execution time than node-move-in, as the number of joining messages are reduced.

When a node joins the network then it sends the gather data to the base station through inter mediate nodes. Therefore an efficient routing algorithm is the must here. QoS-aware routing schemes considers inter cluster communication where few schemes are developed to figure-out intra cluster routing cost. The QoS routing is a difficult and challenging task. In this paper Data Centric Routing is used for intra cluster communication in DCWSN. It considers data reliability and end-to-end path delay. It is evaluated that both the algorithms are outperformed comparatively current algorithms.

This paper is organized as follows. In section 2, literature review for wireless sensor network is discussed. Proposed algorithms are described in section 3. Section 5 describes Simulation results. Conclusion and Future works is described in Section 6.

II. Literature review

The attention of Wireless sensor networks have been gaining rapid increase for past few years. The state-of-the-art of cluster-based architecture has been reported different issues in the network. This section discusses various robust cluster-based schemes.

Leach Mobile (Leach-M) proposes mobility centric protocol for wireless sensor network [16]. Its design consideration is to overcome data packet loss in mobile nodes during node joining. In Leach-M, before data transmission, the new mobile node sends a join request to CH and waits for acknowledgement. The purpose of this joint request is to know whether it is part of the CH or not. If the node receives the acknowledgement within the before expiration of time, then sends the data to CH. Consequently, new node decreases the packet loss and makes successful data communication. If the new node does not receive acknowledgement within defined time, it considers that the node is no more part of the cluster. The computation rounds of Leach-M is $O(n + |V|)$ and communication complexity is $O(n)$.

Leach Mobile Enhancement is a Cluster-based protocol which is suitable for single new mobile node joining the network [17]. The design consideration of the algorithm is to enhance the network lifetime. Each new node looks for CH having low mobility comparatively other node. To measure the low mobility also consider two important parameters those are, remoteness and time. Remoteness has an association with communication link modification rate. Consequently, this consideration forms uniform speed clusters, where CH has minimum speed in the group. The advantage of such groups is that, the average maintenance is small to maintain high spatial dependency. Thus, node movements do not breakage the relationship of MNs with CH. Node elects a lesser mobility factor for cluster head to become member node. The computation rounds of Leach-ME is calculated $O(n)$ and communication complexity is calculated $O(|q|)$.

Another Novel Cluster-Based Architecture and a Routing Protocol for Dynamic Ad-Hoc Radio Networks CBNet(G) [20] is proposed that is mainly designed to support a time and energy efficient, loop-free, on demand routing protocol. In the proposed architecture nodes are capable of performing new nodes joining using node-move-in. New node determines its status according to the recursive structure of CBNet(G) as follows. If there exists CH(s) in the neighbor of new, new becomes an ordinary node and chooses a CH node as new's parent. Else if new finds any cluster forwarding node in its neighbor, new becomes an ordinary node and chooses one of the cluster forwarding nodes as new's parent that has the minimum distance to its CH. Else if new finds any forwarding node in its neighbor, new becomes an ordinary node and chooses one of the forwarding nodes that have the minimum distance to its CH as its parent. Else if new finds any ordinary node in its neighbor who is closer to CH k (k≥ 2), new selects one to be its parent which has the minimum distance to its CH; new becomes an ordinary node and the selected node changes its status to forwarding node. Else if new finds gateway nodes in its neighbor, new selects one as its gateway and new becomes a CH of a new cluster. Else, there are only ordinary nodes in the neighbor of new whose distance to their CHs are already k. New selects one ordinary node as its parent and becomes a CH of a new cluster. The chosen node then changes its status to gateway node and all other nodes that lie on the path from the gateway node to its CH change their status to cluster forwarding node [20]. The computation rounds of CBNet(G) is calculated $O(q + k)$ where q denotes neighboring node of new node and k denotes maximum radius and communication complexity is $O(r + l)$ where r is number of intra-cluster nodes and l is size of inter-cluster.

Mobility based Cluster (MBC) protocol is suggested that is well suitable when new nodes are mobile [10]. MBC takes into consideration of two important parameters for CH election in new nodes which are remaining energy and expected connection time. For CH election considers more remaining energy of the nodes. Moreover, to overcome the problem of packet loss introduces a solution of expected connection time between CH and MNs.

DCWSN architecture is address in [12-13]. The architecture is maintained using efficient maintenance schemes, namely node-move-in and node-move-out. Node-move-in addresses single new node joining while node-move-out addresses single node leaving. Initially new node searches CH then gateway node and at last member node in neighboring node. If neighboring node is a CH then new node becomes MN. Elseif, neighboring node of new node is gateway node then new node become CH. Else neighboring node is member node then neighbor node change its status from member node to gateway node and new node becomes CH. To get single node joining requires $O(q)$ rounds.

III. Proposed algorithm

In this section two algorithms are presented, namely Multi nodes joining with complete 1-hop neighbor information network and Data Centric Routing. Few considerations are formed for the development of the algorithms which are discussed in DCWSN Model in section *A*. Section *B* discusses Multi nodes joining with complete 1-hop neighbor information network. Data Centric Routing is discussed in section *C*.

A. DCWSN Model

The following are the considerations for multi nodes joining to DCWSN.

Considerations for multi nodes joining to DCWSN model are discussed in this section. In the proposed model, all the nodes in the network are considered to be static, where each node has unique ID and knows the maximum network size. The maximum network size is also referred as upper bound number of nodes in the network in this thesis. A root node is considered in the network, where the status of the root node is considered to be CH. It is also considered that the root node has the entire network information such as status and ID of each node and information about all the edges in the network.

The network is considered as single channeled. In the network, nodes use different channels for data messaging and control messaging. Thus, nodes use one channel for data transmission while another channel for network maintenance. When more than one node transmits at same time, the message is discarded because of collision. Thus the message is not recognizable by receiving nodes. However, if there is no collision, message delivery is guaranteed for the sender node. The network has no-collision-detection model. Therefore, collision is not recognizable by nodes. Every node is equipped only one transceiver, so that a node cannot perform transmitting and receiving operations at the same time. The communication range of each node is similar except the root node. In this article, communication range is considered as the maximum distance between two nodes where these two nodes can receive the transmitted data correctly.

Moreover in the network, node(s) that is within the communication range of new node(s) is represented by "the existing neighbor nodes in DCWSN". In the proposed network model, a complete 1-hop neighbor information network is defined as a network where each node has all its 1-hop neighbors' information, such as their status and ID.

B. *PROPESE MULTI NODES JOINING WITH COMPLETE 1-HOP NEIGHBOR INFORMATION (MNJ-CN)*

This section proposes Multiple nodes joining with complete 1-hop neighbor information (MNJ-CN) algorithm to join nodes the network as describe in Figure 1.

New nodes are randomly deployed in the network field in order join the network. This is due to increase the network coverage of the network. Each new node competes the channel to send *Hello* message and successfully join the network. Thus a node among all new nodes that successfully sends *Hello* becomes the *winner*. As few new nodes might be in the range of *winner* node, therefore these new nodes also hear the *winner* node. Consequently, these new nodes stop competing the channel until the *winner* node joins the network. This is because that the nodes do not interrupt the winner node joining procedure.

Upon sending *Hello* message to neighbor nodes, neighbor nodes respond with *Hi* message. Neighbor nodes are DCWSN nodes those are within the range of the winner node.

When the *winner* node receives the *Hi* message, initiates joining the network by sending the joining message to the existing neighbor nodes. When the existing neighbor node in DCWSN *q* receives the joining message from the *winner* node for the very first time, responds the ACK where ACK message consists of *q*'s ID and status.

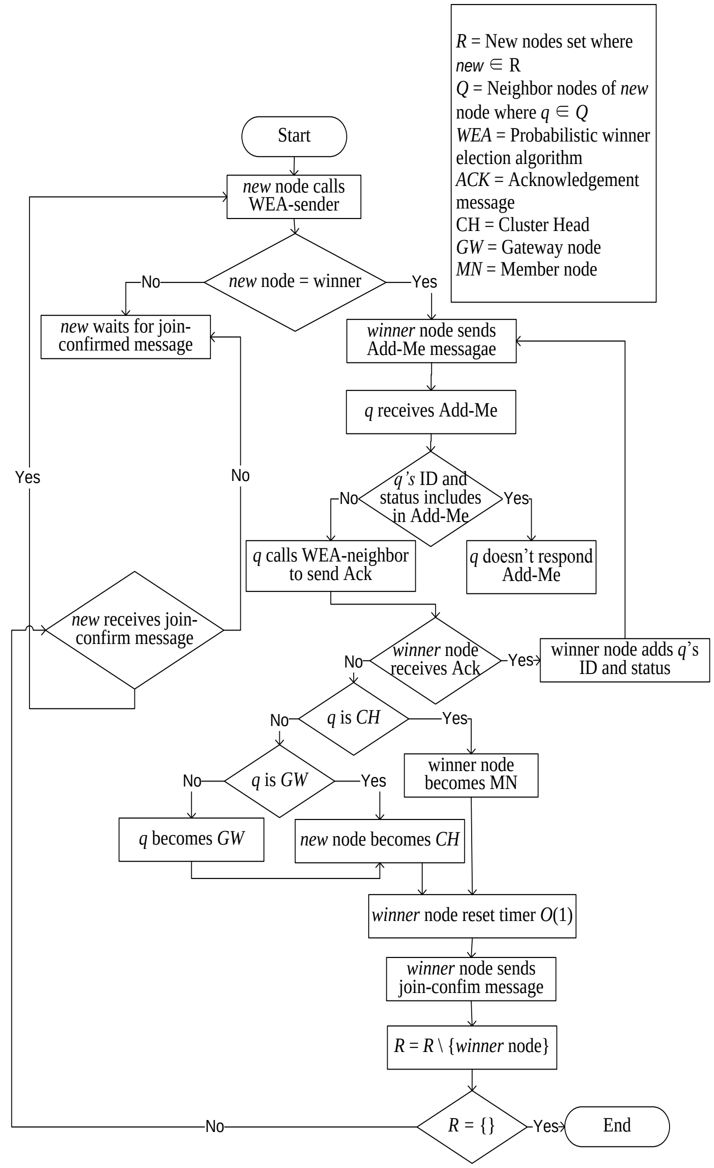


Figure 1: Flowchart of the algorithm

Upon receiving *q*'s ACK to *winner* node, *winner* node saves *q*'s ID and status and execute the next step immediately without any delay. The *winner* node resends the joining request to the network in the next step. When the message is received to *q*, looks its own ID in the message. When *q* finds its own ID, discards the message and do not send the ACK again. Thus the number of messages in new nodes joining are minimized which eventually makes the algorithm efficient in terms of faster joining new nodes to the network. When *winner* node doesn't get the ACK message from *Q*1, stops sending the message anymore and execute the node joining procedure which is also defined below. Figure 1 illustrates the joining process of new nodes. .

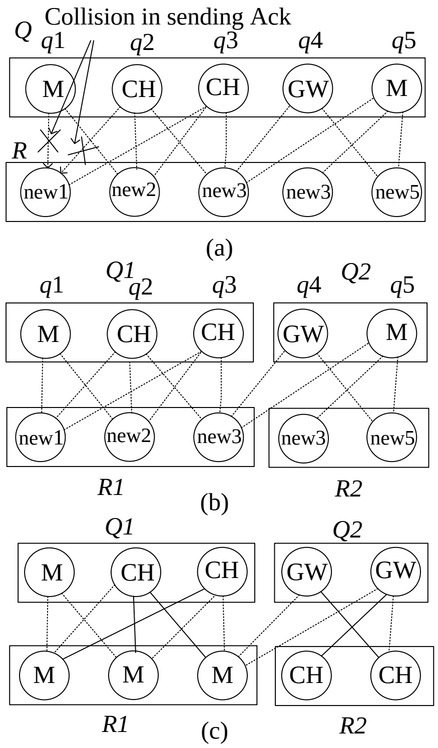


Figure 1 new nodes joining using MNJ-CN

Let *new*1 sends the Add-Request to neighbor node *Q*1 and respectively *Q*1 responds the ACK message to *new*1. Since *Q*1 is within the range of *R*1 where *R1* = {*new*1, *new*2, *new*3} in Figure 1.b. Therefore, the ACK also receives to *new*2 and *new*3 too. Thus, eventually *R*1 also stores the ACK message. Next, *new*1 sends the request message which also consists of IDs of the neighbor nodes in DCWSN which is already received IDs. When the request message is received by neighbor node *q*, the node searches its own ID in the message. However *q* do not respond ACK when it finds its own ID in the message. When *q*'s ID is missing, sends ACK to the sender node. When *new1* does not obtain the message from *q* while the timer expires, it assures that *new1* has already gathered entire existing neighbor information of DCWSN *Q*1. Thus, *new*1 joins the network and decides its status that is defined below.

The *new*1 looks for an appropriate node in order to join the network. When *Q* includes CH (*q*), *winner* node becomes member node of *q* where, *winner node*'s state becomes member node (MN), *winner node's* parent becomes *q*. When *Q* does not include CH node and includes GW (*q*) then *winner node's* state becomes CH, *winner node's* parent becomes *q*. On the other hand when *Q* include MN (*q*), *q* changes the status from MN to *GW*. When *winner* node receives response message about status change of the *MN* (*q*) to GW, *winner* node becomes CH, wherein *winner node*'s state becomes CH, *winner* node's parent becomes *q* (Figure 1.c). When $|Q| \geq 2$ and have the same status then *winner* node chooses a node with lesser ID as its own parent node. Once the *winner* node (*new*1) joins the network, it sends join confirmation message to neighbor nodes.

When *R* receives the join confirmation message from *new*1, *R* competes the channel to become *winner* node. Let *new*2 be a *winner* node among *R* where *R* = {*new*2, *new*3, *new*4, *new*5}. Now *new*2 (Figure 1.b) sends an Add-Request which include known IDs of *Q*1, to the existing neighbor nodes in DCWSN *Q*1. However, this time *Q*1 does not send the ACK message, as *Q*1 checks their IDs and finds its own ID in the Add-Request. Thus *new*2 node timer is expired to hear a new message from the neighbor node. Thus *new*2 eventually joins the network as per aforementioned criterion. Thus, the number of messages are minimized when neighbor nodes do not send the ACK over and over again. When *new*2 joins the network, it sends Join-confirmed to *R*.

Similarly, let *new*3 be a *winner* node in *R*. *New*3 sends the joining request which also includes known IDs of *Q*1, to neighbor nodes. Now, *new*3 receives ACK from neighbor *Q*2 (Figure 1.c) where *Q*2 = {*q*4, *q*5}, *q*2 and *q*3 do not send the ACK as these nodes IDs are included in the joining message. Now *new*3 also joins the network as per aforementioned criterion.

To sum-up, minimize number of messages to form complete 1-hop neighbor information network, each new node searches missing neighbor nodes by sending neighbor nodes IDs which is already received. All the existing neighbor nodes in DCWSN check the IDs with their own ID upon receiving the Add-Request. When any neighbor node in DCWSN finds its information missing, the node send their information to new node. The *new* node stops searching when it does not receive any message from the existing neighbor nodes in DCWSN and join the network.

Subsequently, remaining nodes joined the network as per aforementioned method.

Theorem 1*: Let, DCWSN be a cluster-based network of flat network G. G is organized with complete* 1-hop neighbor information *network. Q be the set of neighbor of new nodes R. When G is organized with complete 1-hop neighbor information network, MNJ-CN can be achieved in* $O|R|((log|R|logN) + O(|Q|logQ))$ *rounds.*

Proof*:* As presented in theorem 1, new nodes (*R*) calls WEA-sender to become winner that is achieved in $O(log|R|logN)$ rounds. *winner* node sends a joining message to the existing neighbor nodes *Q*. When *Q* receive a joining message then it sends a respond message. In order to send the respond message without collision the neighbor nodes in DCWSN calls WEA-neighbor that requires $O(logQ)$ rounds. WEA-neighbor enables to reduce the number of new nodes that attempt the channel. Hence, after $O(logQ)$ rounds, the number of sender node becomes 1. The $O(logQ)$ implies of dividing *Q* set to halves until $log(Q)$ times, thus achieve $|Q| = 1$. Winner node gathers complete 1-hop neighbor information network to decide its own joining status. Since single winner is achieved in $O(logQ)$ rounds, thus, to gather all the neighbor nodes information in order to achieve complete 1-hop neighbor information requires $O(QlogQ)$ rounds. Hence joining a winner node to DCWSN is achieved in $O(log|R|logN) + O(|Q|logQ)$ rounds. Thus joining multi nodes using MNJ-CN requires $O(|R|((log|R|logN) + O(|Q|logQ))$ rounds.

Next section provides proposed protocol for routing algorithms.

IV. DATA CENTRIC ROUTING ALGORITHM

The proposed routing algorithm Data Centric Routing is a cross layered approach uses different modules for various tasks. The following sub-sections discuss the other modules of the proposed scheme. The Data Centric Algorithm introduced in [21] and is reused in DCWSN for efficient data routing. The Data Centric Algorithm addresses delay and reliably in the network. Following is a brief overview of Data Centric Algorithm.

A. Delay-awar module

The task of delay-aware module is to select best possible route to send message. At any node n_i, when a message P is received, the Delay-Aware Algorithm, looks at the routing table RT and selects those neighbor nodes belongs to routing table whose link quality $LQ_{i,j}$ between nodes n_i, and n_j, is higher or equal to the re- quired link quality LQ_{req} and store them in NN_{LQ}. The message is P is dropped in case of no such node. In case of non-empty NN_{LQ}, it chooses only those neighbors nodes belongs to NN_{LQ}, whose end-to-end path delay $PD_{i,j}$, from node ni to BC through node n_j, is less than or equal to the required delay D_{req} and store them into NN_D. In case of null NN_D, the message P is dropped, in case single entry in NN_D, that single neighbor node is selected as desired next hop DNH. On the other hand, in case of more than one entry in NN_D, temperature-aware module is called with message P and NN_D as inputs [21].

Delay measures the waiting time due to queuing, processing, propagation and transmission at any $node_i$ Queuing and transmission are the dominating factors that causes delay in message transmission.

$$QD_{ni} = \alpha QDni + (1 - \alpha)QD_{ni} \tag{1}$$

Temperature is evaluated using equation (2), (3), and 4 respectively.

$$PDi,j = PD_{ij} + ND_{ni} \tag{2}$$

$$PRij = PR_{ij} * LR_{ij} \tag{3}$$

$$PT_{ij} = PT_{ij} + NT_{ni} \tag{4}$$

The transmission delay TD_{ij} of the link $L_{i,j}$ between nodes n_i and n_j is the transmission interval between the time when a message enters the MAC layer to the time when it is either successfully transmitted or dropped. NP is the number of messages transmitted in time specific time window. At any node n_i the node delay is given by equation 5.

$$NDni = QD_{ni} + TD_{i,j} \tag{5}$$

B. Reliability-aware module

The reliability-aware module of the proposed algorithm is responsible to choose efficient route for the message. Once a data packet P is received at reliability aware module, the Reliability Aware Algorithm given in Algorithm, searches the routing table RT and the desired next hop DNH node is selected in the same manner as in delay-aware module except that the decision is made based on the end-to-end path reliability $PR_{i,j}$, from node n_i to BC through node n_j. The reliability link between two neighbors can be calculated by equation 6 [21].

$$P_{average} = N_{ack}/N_{trans} \tag{6}$$

C. Energy estimator

Sensor nodes share their remaining energy to their neighbor nodes when sends message to destination node. Each node stores 1-hop neighbor nodes energy information in their routing table. As a result the node avoid to send the message to a node that remaining energy is lesser than the minimum threshold level. Consequently the network lifetime is enhanced along with reliable data transmission the network [21].

V. SIMULATION RESULTS

To measure the performance of the algorithms, simulation is conducted using MATLAB. The simulation results are analyzed to evaluate the efficiency of MNJ-CN and Data Centric algorithms. Number of messages and execution time are the two parameters used to analyze the performance of the proposed schemes which are described below.

A Number of messages

The number of exchange messages such as, Hello, AddMe from a sender node and acknowledgements messages from responder node during node joining and leaving are calculated as a number of messages. The meanings of sender and responder are very simple. The node that transmits the data is known as sender and the node that receives the data is known as responder.

B Execution time

Execution time describes the total time duration that an algorithm needs to execute. The lesser the execution time of the algorithm, the lighter the algorithm.

The simulations of the proposed schemes are performed 10 times and the average of those 10 results is taken. For the purpose of simulation, nodes are deployed in 500m x 500m field. New nodes are deployed randomly. Moreover, the transmission range of each sensor node is set to be 70 m. A node is elected as a root node and set its status as a cluster head, WEAs (WEA-sender, WEA-neighbor) and MNJ-CN are executed, respectively.

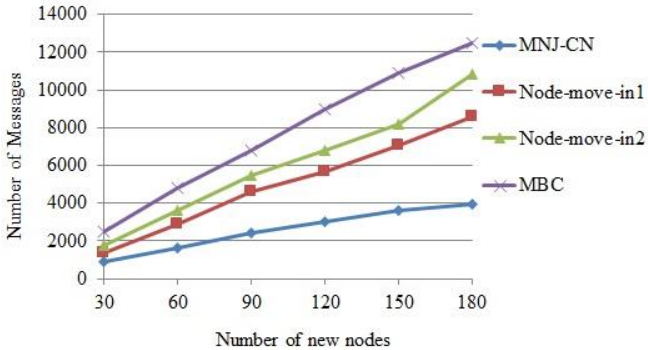


Figure 1: Describes number of new nodes versus number messages.

Figure 1 describes the performance of MNJ-CN with existing algorithms; node-move-in1, node-move-in2 and MBC in terms of number of messages. The result shows that when 30 nodes are joined through MNJ-CN, node-move-in1 [12-13], node-move-in2 [20] and MBC [10] algorithms, they require 930, 1420, 1800 and 2540 number of messages, respectively. When number of new nodes are exceeded to 180, they require 3919, 8556, 10800, 12500 number of messages, respectively. Here, it can be seen that for any number of new nodes, MNJ-CN requires less number of messages as compared to the existing algorithms in joining. This shows that MNJ-CN has better performance than other three algorithms.

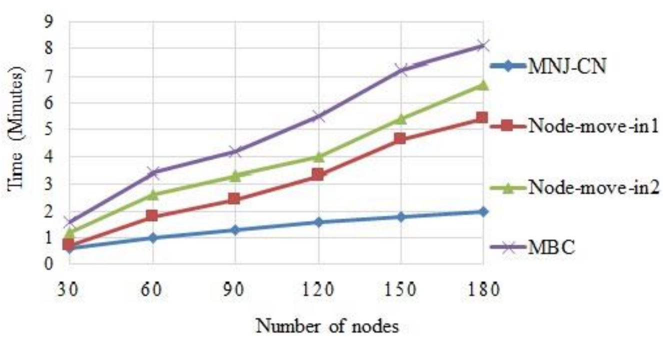


Figure 2: Performance Evaluation for MNJ-CN

As indicated in Figure 2, with the increasing number of new nodes, the time taken for MNJ-CN, node-move-in1, node-move-in2 and MBC are also increases. However, it can be seen that the joining time for MNJ-CN require lesser time than Node-move-in1 [12-13], Node-move-in2 [20] and MBC [10]. This is because nodes are already gathered information of their neighbor nodes.

VI. CONCLUSION AND FUTURE WORKS

In this paper MNJ-CN and Data Centric algorithms are presented. MNJ-CN successfully joins the node and also minimizes number of messages in joining process.

It is observed from the simulation experiment that MNJ-CN also requires lesser number of messages and lesser execution time compare to current algorithms which outperforms than present algorithms.

In future, the research can be focused on the following directions: firstly, fault-tolerance and network robustness would be designed. Secondly, security mechanism would be introduced which would secure the network from malicious nodes. Thirdly, an efficient routing protocol could be introduced to reduce end to end delay. Fourthly, a dynamic clustering would be examined in diverse areas like energy efficiency, load balancing, network robustness and degree based schemes. Fifthly, the concept of multi nodes joining the network would be proposed for real time applications in order to enhance its effectiveness.

REFERENCES

[1] K. M. Alzoubi, et al., "New distributed algorithm for connected dominating set in wireless ad hoc networks," 2002, pp. 3849-3855.

[2] A. D. Amis and R. Prakash, "Load-balancing clusters in wireless ad hoc networks," in Application-Specific Systems and Software Engineering Technology, 2000. Proceedings. 3rd IEEE Symposium on, 2000, pp. 25-32.

[3] X. Liu, "A survey on clustering routing protocols in wireless sensor networks," Sensors, vol. 12, pp. 11113-11153, 2012.

[4] U. Prathap, et al., "Wireless Sensor Networks Applications and Routing Protocols: Survey and Research Challenges," in Cloud and Services Computing (ISCOS), 2012 International Symposium on, 2012, pp. 49-56.

[5] N. Vlajic and D. Xia, "Wireless sensor networks: to cluster or not to cluster?," in Proceedings of the 2006 International Symposium on on World of Wireless, Mobile and Multimedia Networks, 2006, pp. 258-268.

[6] A. Zeb, et al., "Multi-nodes joining for dynamic cluster-based Wireless Sensor Network," in Informatics, Electronics & Vision (ICIEV), 2014 International Conference on, 2014, pp. 1-6.

[7] J. Y. Yu and P. H. J. Chong, "A survey of clustering schemes for mobile ad hoc networks," IEEE Communications Surveys and Tutorials, vol. 7, pp. 32-48, 2005.

[8] N. S. Yadav, et al., "A Low Control Overhead Cluster Maintenance Scheme for Mobile Ad hoc NETworks (MANETs)," ACEEE International Journal on Network Security, vol. 1, 2010.

[9] M. Younis and K. Akkaya, "Strategies and techniques for node placement in wireless sensor networks: A survey," Ad Hoc Networks, vol. 6, pp. 621-655, 2008.

[10] S. Deng, J. Li, and L. Shen, "Mobility-based clustering protocol for wireless sensor networks with mobile nodes," Wireless Sensor Systems, IET, vol. 1, pp. 39-47, 2011.

[11] R. A. Uthra and S. Raja, "QoS routing in wireless sensor networks—A survey," ACM Computing Surveys (CSUR), vol. 45, p. 9, 2012.

[12] J. Uchida, et al., "Construction and maintenance of a novel cluster-based architecture for ad hoc sensor networks," Journal of Ad Hoc and Sensor Wireless Networks, vol. 6, pp. 1-31, 2008.

[13] A. K. M. M. Islam, et al., "A Better Dynamic Cluster-Based Structure of Wireless Sensor Network for Efficient Routing," International Journal of Innovative Computing Information and Control, vol. 8, pp. 6747-6760, Oct 2012.

[14] G. S. Kumar, et al., "Mobility metric based leach-mobile protocol," in Advanced Computing and Communications, 2008. ADCOM 2008. 16th International Conference on, 2008, pp. 248-253.

[15] S. A. Awwad, et al., "Cluster based routing protocol for mobile nodes in wireless sensor network," Wireless Personal Communications, vol. 61, pp. 251-281, 2011.

[16] D.-S. Kim and Y.-J. Chung, "Self-organization routing protocol supporting mobile nodes for wireless sensor network," in Computer and Computational Sciences, 2006. IMSCCS'06. First International Multi-Symposiums on, 2006, pp. 622-626.
[17] G. S. Kumar, et al., "Routing protocol enhancement for handling node mobility in wireless sensor networks," in TENCON 2008-2008 IEEE Region 10 Conference, 2008, pp. 1-6.
[18] S. Deng, et al., "Mobility-based clustering protocol for wireless sensor networks with mobile nodes," Wireless Sensor Systems, IET, vol. 1, pp. 39-47, 2011.
[19] O. Younis and S. Fahmy, "HEED: a hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks," Mobile Computing, IEEE Transactions on, vol. 3, pp. 366-379, 2004.
[20] A. M. Islam, Y. Katayama, W. Chen, and W. Wada, "A novel cluster-based architecture and routing protocols for dynamic ad-hoc radio networks," Journal of Electrical Engineering, The Institution of Engineers, Bangladesh, 2006.
[21] J. Iqbal, A. Khan, A. Abdullah, "Data-centric routing for intra wireless body sensor networks" Journal of Medical Systems 39.9 (2015): 1-13

Bangla Word Sense Disambiguation System using Dictionary Based Approach

Afsana Haque and Mohammed Moshiul Hoque
Department of Computer Science & Engineering
Chittagong University of Engineering & Technology
Chittagong, Bangladesh
Email: afsanahaque22@gmail.com; moshiulh@yahoo.com

***Abstract*: Bangla Word sense disambiguation (WSD) is the task of selecting the appropriate senses of a Bangla word (i.e. meaning) in a given Bangla sentence when the word has multiple meanings. This thesis work focuses on the issues in designing a Bangla Word Sense Disambiguation System to remove the lexical semantic ambiguity by using dictionary based approach. This system works with two major steps: parsing and detection. We propose algorithms for these two major steps and remove the ambiguity for three types of ambiguous word as noun, adjective and verb in the Bangla sentences. The evaluation result reveals that the proposed word sense disambiguation system can disambiguate the ambiguous Bangla sentences with 82.40% accuracy for some selected sentences.**

***Keywords*: Word sense disambiguation; dictionary based approach; context sensitive grammar; parsing**

I. INTRODUCTION

Many words have multiple meanings, depending on the context in which they are used. Word sense disambiguation (WSD) is an ambiguity removal technique of determining the correct sense of a word (i.e. meaning) in a given sentence, when the word has multiple meanings. Basically WSD requires a set of meanings for each word to be disambiguated and a means to choose the correct one from that set. Many factors contribute to the difficulty of WSD including words with multiple meanings, sentences with multiple grammatical structures and so on.

For humans, resolving ambiguity is a routine task that hardly requires conscious effort. In addition to that need to a deep understanding of language and its use. Humans possess a broad and conscious understanding of the real world and this equips them with the knowledge that is relevant to make sense disambiguation decisions effortlessly, in most cases. However, creating extensive knowledge-bases which can be used by computers to 'understand' the world and reason about word meanings accordingly, is still an unaccomplished goal of Artificial Intelligence (AI). This makes the creation of programs that disambiguate a word, one of the most challenging tasks in natural language processing.

WSD is regarded as an important research problem and is assumed to be helpful for machine translation (MT). There is no doubt that Bangla WSD is essential for the proper translation of polysemous Bangla words when we will try to translate a Bangla sentence into another natural language. So by implementing Bangla WSD technique we can spread Bangla all over the world through then translation of Bangla language properly. Besides this the system is also important for human computer interaction (HCI). As HCI is focusing on the interfaces between people (users) and computers, this system is motivating people to learn the Bangla language by themselves using computer and to prompt the development of technologies. Moreover, it helps to motivate a student to learn Bangla language properly in the absence of their teacher. Nowadays we have many Bangla websites and blogs but those are not enough to satisfy our need for knowledge. But a Bangla WSD system can help all the people of Bangladesh to gather meaningful information.

WSD system easily can be developed by using dictionary based approach. The dictionary based word sense disambiguation technique is one of that kinds of knowledge based approach which relies on machine readable dictionaries (MRD). The dictionary contains the lexicons of Bangla words. It also serves the storage of ambiguous Bangla words and their corresponding meanings. In this approach, at first all of the sense definitions of each word in a Bangla input text are retrieved from this MRD. Then compare each sense with the word which has more than one sense. Finally the word is selected which is the most overlapped with the dictionary definitions. Here errors can be adjusted safely through changes to this user dictionary. So it can be quite costly and time consuming to implement and maintain.

II. RELATED WORK

Research on Bangla word sense disambiguation technique is comparatively rare. It is now in rudimentary stage. There are limited number of research work is done under dictionary based approach for this Bangla WSD system. A recent work has been done using existing lexical knowledge sources such as WordNet [1].

Previous work [2] describes the theoretic Bangla Word Sense Disambiguation technique using Lexeme List. By using supervised algorithm they only find the ambiguous word but they don't show exact meaning of the sentence because of morphological complexity. Pal et al. [3] proposed a Naive Bayes probabilistic model to classify the sentences and the target word they used in their experiment was a noun but they didn't use this method for a verb. Pal et al. [4] proposed an approach to generate their output with the help of a Bengali lexical dictionary. They lemmatized the test data obtained around 85% accurate results. A k-Nearest Neighbor (k- NN) algorithm is proposed in [5] for their research work and achieved an accuracy of over 71%.D. McCarthy et al.[6] implemented the Yarowsky algorithm [7] as a solution and analyzed its behavior with respect to program parameters. Previous work [8] describes induction and related approaches and [9] describe an existing measure for semantic relatedness between two lexically expressed concepts of Hindi WordNet.

Some authors are proposed the dictionary based approaches in [10,11,12,13].

III. PROPOSED BANGLA WSD SYSTEM

The schematic representation of proposed dictionary based Bangla WSD system is illustrated in Figure-3.1.

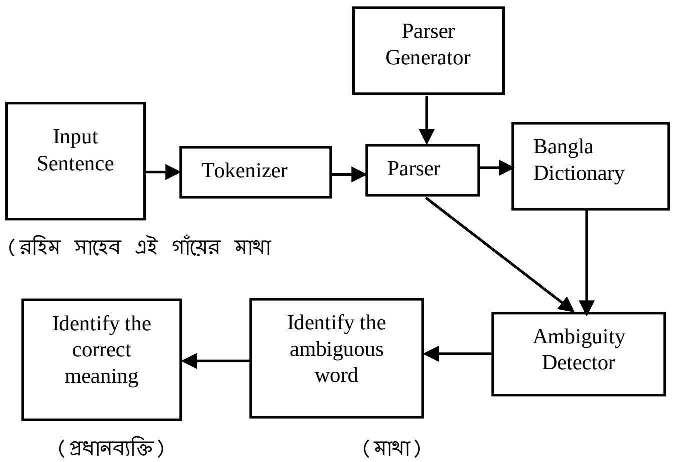


Figure 1: Proposed WSD System.

A. *Tokenizer*

Input sentence will first enter in Tokenizer. Tokenizer is the program module that accepts a sentence to be parsed as an unbroken string, breaks into individual words called Tokens [14]. Tokens are stored in the list for further access. These tokens are then entered into the parser to generate parse tree. For example, if we consider
Input sentence: "রহিমসাহেবেরমাথাব্যথাকরছে"
Output of tokenizer: ("রহিম", "সাহেব", "এর", "মাথা", "ব্যথা" , "কর", "এছে").

The result of tokens is used as input of parsing to the structure of input Bangla sentence and the tokens are checked in the dictionary for validity.

B. *Bangla Dictionary*

A Bangla dictionary is used to store the words and also to store the corresponding Bangla meanings if any Bangla word has several senses. It is connected with the parser. When it is called, it serves various Bangla words to the parser for the verification of the parse tree. Then the ambiguity detector will check and detect the ambiguous word if it has several meanings by using this dictionary. It also serves the actual meaning to the ambiguity detector. A screen shot of used Bangla dictionary is given below in table 1 and 2. Table1 is showing the Bangla words, their type and their features for our input Bangla sentences. And table 2 is showing only the ambiguous word and their possible corresponding actual meanings.

TABLE I.TYPICAL ENTRIES OF SOME BANGLA WORDS IN DICTIONARY

Bangla Word	**Type**	**Feature**
রহিম	Noun	Human, Pers2
সাহেব	Noun	Human, Pers2
আমি	Pronoun	Human, Pers1
সে	Pronoun	Human, Pers3
মাথা	Noun	Organ
কান	Noun	Organ
ভাল	Adjective	Qualitative
গ্রাম	Noun	Place
ছাড়া	Verb	Transitive, Intransitive
অভ্যাস	Noun	Distress
আবশ্যক	Adverb	Positive
নাই	Indeclinable	Negative
তাড়াতাড়ি	Adverb	Manner

TABLE II. TYPICAL ENTRIES OF CORRESPONDING MEANINGS OF SOME BANGLA WORDS IN DICTIONARY

Bangla word	**Corresponding word**	**Meaning**
মাথা	গাঁ, গ্রাম	প্রধান ব্যক্তি বা মোড়ল
মাথা	ব্যথা, ধরা	শারীরিক কষ্ট
মাথা	কাটা	লজ্জা পাওয়া
মাথা	চোখের	অন্ধ হওয়া
মাথা	খারাপ	বিরক্ত করা
পাকা	বাড়ি	ইট, সিমেন্টের নির্মিত দালান
পাকা	আম	পরিপক্ক
পাকা	চুল	সাদা হয়ে যাওয়া
পাকা	চাকরি	স্থায়ী হওয়া
পাকা	খবর	সঠিক
ছাড়া	হাত	নিরাশ হওয়া
ছাড়া	চাকরি	শেষ ভরসা বা অবলম্বন
ছেড়ে	দিয়েছে	মুক্ত হওয়া
ছেড়ে	আশা	আশাহত হওয়া
ছেড়ে	চাকরি	ত্যাগ করা
মুখ	রাখল	গৌরব বাড়ানো
মুখ	চেয়ে	নির্ভর করা
মুখ	চুনকালি	লজ্জিত করা
মুখ	উপরেই	অপমান করা
মুখ	ভার	রাগ করা

C. *Parser Generator*

Tokens are now used as the input of the parser generator to generate the parse tree. The most common way to represent grammar is as a set of production rules which says how the parts of speech can put together to make grammatical or well-formed sentences [14]. In this thesis, we have used CSG rules to make the Bangla parse trees. Basically ambiguous word can be three types as noun, adjective and verb. Considering this, we have merged all those rules for each type of individual sentences and use only three or four rules for our parsing. As our selected Bangla sentences are simple but some are too long in size. We have taken help from the book [15]. In this book, the parse tree generation rules are described in details and these rules are also applied on all types of Bangla

sentences such as simple, complex and compound with examples.
Here the list of some Bangla CSG rules is given in table 3.

TABLE III. LIST OF BANGLA CSG RULES

Rule No.	Bangla CSG Rules
1.	S→NP VP | VP NP
2.	NP→N | PN | ADJ
3.	NP→ N(BIV)N | N(BIV)(Adj) | PN(Adj)
4.	NP→N(N)(DET)(BIV)(Con) |N(Con)(Adj)
5.	NP→N AdjAdj | PN PN N(Con)
6.	NP→N PN N(Con)
7.	VF→VR | (Adv) (NP) VF | VR(BIV)(Con)(ind)
8.	VP→NP (Adv) VF | V(ind)(ind)
9.	VF→(ind) V (BIV)
10.	NP→ null
11.	VP→ null
12.	N→রহিম, সাহেব, গ্রাম, মাথা, কান, মুখ....
13.	PN→এই, আমার, সে, আমি, কোন, তার....
14.	V→কর, কিনব,পড়, তোলা, দাও....
15.	Adj→অনেক, অসহনীয়,খারাপ, দিক, নতুন....
16.	Adv→আবশ্যক, এসে, কথায়, করলে, চলে....
17.	Con→লাম, এছে, ইস, আস, আছ....
18.	BIV→এ, য়, তে, কে...
19.	ind→না, নাই, নেই....
20.	DET→টি

[Abbreviations: S: Sentence, NP: Noun phrase, N: Noun, PN: Pronoun, VP: Verb phrase, VF: Verb form, V: Verb, Adj: Adjective, Adv: Adverb,BIV: Bivokti (inflection), Con: Concord, ind: indeclinable, DET: Determinant]

D. Parser

Parser is a procedural component and is nothing but a computer program. It is the process of taking tokens of input sentence and producing a parse tree according to CSG rules. The result of parser is a parse tree. Basically a parser is the component of parsing. At parsing stage, the tokens are checked into the dictionary for the verification. And then a set of rules will be used in our framework. With variation of rules, parse tree will be also different from each other. For example if we consider,

Input sentence: "রহিম সাহেবের মাথা ব্যথা করছে"

Parser output: S[NP[[N রহিম][N সাহেব][Con এর]][VP[NP[[N মাথা] [N ব্যথা]]VF[[V কর][Con এছে]]]]]

The basic structure of the parse tree of corresponding example is given in figure 2:

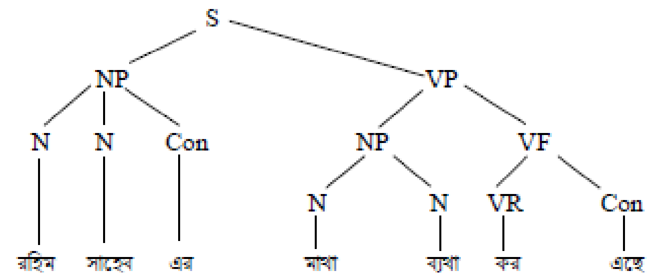


Figure 2: Parse tree of "রহিম সাহেবের মাথা ব্যথা করছে"

1) Parsing Algorithm

There are two ways of parsing a sentence. Top-down and bottom-up parsing. Here the input sentence is parsed by using parsing methodology through the following steps:
Step 1: The input of parser is the output of the tokenizer. Tokens are stored in a stack for further access. For example: if the input sentence is "রহিমসাহেবেরমাথাব্যথাকরছে" then the tokens will be ("রহিম", "সাহেব", "এর", "মাথা", "ব্যথা" , "কর", "এছে"). These tokens will will store in a stack.
Step 2: The tokens are then checked the lexicon for the validity. For example: the tokens ("রহিম", "সাহেব", "এর", "মাথা", "ব্যথা" , "কর", "এছে") will be enter into the bangla dictionary to find its validity.
Step 3: The tokens are matched with the grammar rules. If a rule whose right hand side matches with a token, then the token is assigned with appropriate parts of speech. For example: N→রহিম, N→সাহেব, Con→এরwill produce a partial structure.
Step 4: Starting from left to right hand side of token list, check every rule whose right hand side will match one or more
of the parts of speech. If a right hand side of a rule matches with appropriate parts of speech, then we have to select that rule.
Step 5: Repeat step 4, until no more words to generate.
Step 6: If there are no more words to process, then generate structure of the sentence in a list. For example: the parser output of the sentence "রহিমসাহেবেরমাথাব্যথাকরছে" is given below:
S[NP[[N রহিম][N সাহেব][Con এর]][VP[NP[[N মাথা][N ব্যথা]] VF[[Vকর][Conএছে]]]]]

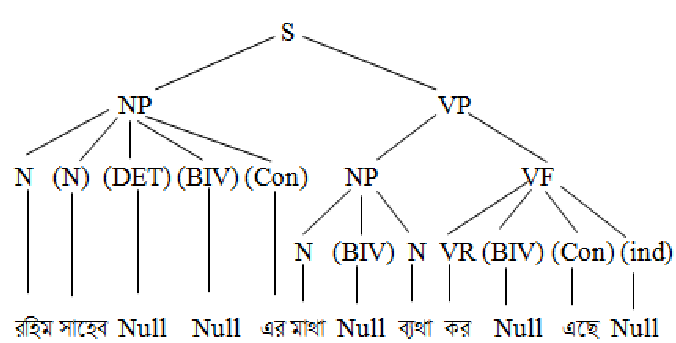


Figure 3: Structural representation of "রহিম সাহেবের মাথা ব্যথা করছে"

E. Ambiguity Detector

Ambiguity detection is the procedure of detecting the ambiguous word in a sentence by using the parse tree. In our system, an ambiguity detector is connected with the Bangla dictionary. And with the help of this Bangla dictionary, it will detect the ambiguous Bangla word. After detecting the ambiguous Bangla word ambiguity detector will try to find out the actual meaning we have stored in the dictionary. By comparing with the corresponding words ambiguity detector will be able to find out the actual sense or meaning.

For example if we consider,
Input sentence: "রহিম সাহেবের মাথা ব্যথা করছে"
Output ambiguous word: মাথা
Output actual meaning: শারীরিক অসুস্থতা

1) Ambiguity Detection Algorithm

The ambiguous bangla word is detected through some steps as given below:

Step 1: At first the ambiguity detector will gather knowledge from the parser and the Bangla dictionary.

Step 2: By using these knowledge it will check each words from the parser with the dictionary. When it get any information that any word has several meanings that are stored in the dictionary, it will notify and detect that word as an ambiguous word. As we have considered two Bangla sentences, we can able to get the Bangla word "মাথা" in both sentences. In first example it is used as "মাথা ব্যথা" which indicates bodily distress and in second example it is used as "গ্রামের মাথা" which indicates head of the village or an important person. As we get two separate meanings for the same word "মাথা", we can say that this "মাথা"word creates ambiguity in the Bangla sentence.

Step 3: When the ambiguity detector can able to detect the ambiguous word it will find the corresponding bangla word which is attachted with the ambiguous word and matched with the dictionary. For our considered example when the ambiguity detector gets "মাথা + ব্যথা" i.e. "মাথা" is attached with the corresponding bangla word "ব্যথা", it shows the actual meaning as "শারীরিক অসুস্থতা". And when it gets "গ্রাম+মাথা" i.e. "মাথা" is attached with the corresponding bangla word "গ্রাম", it shows the actual meaning as "প্রধান ব্যক্তি".

Step 4: If that ambiguous bangla word get matched with the corresponding bangla word it has, it will show the actual meaning.

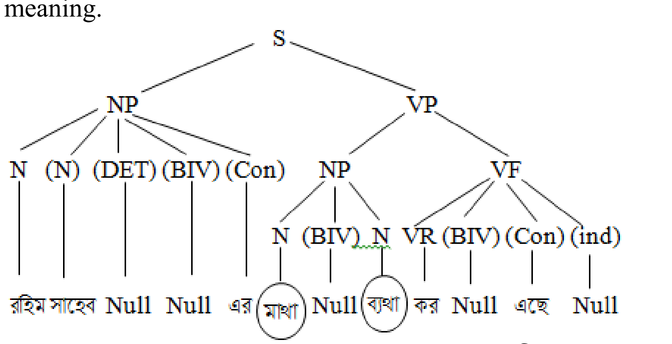


Figure 4: Ambiguity detection for the parse tree of"রহিম সাহেবের মাথা ব্যথা করছে"

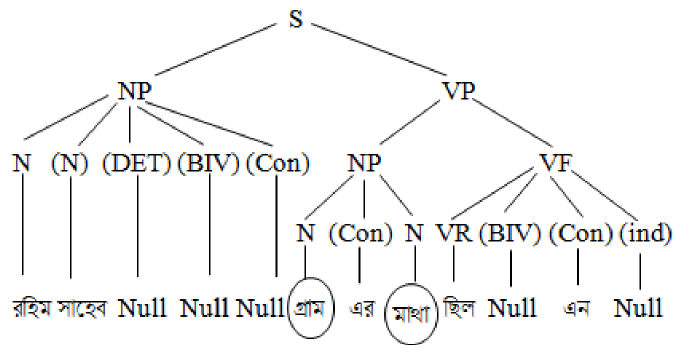


Figure 5: Ambiguity detection for the parse tree of"রহিম সাহেব গ্রামের মাথা ছিলেন"

Step 5: Finally, we get the ambiguous bangla word and its corresponding actual bangla meaningaccording to our input bangla sentence.

IV. Resultant Output

Development of word sense disambiguation system is a challenging task. We have tried our best to develop the ambiguity removal technique efficiently. Basically the ambiguous word can be three in types. It can be a noun or an adjective or a verb. Now the resultant output of our proposed Bangla WSD system for these three types of ambiguous word is representing separately in figure 6, 7 and 8.

The snapshot for the output where the ambiguous word is used as a noun is given below:

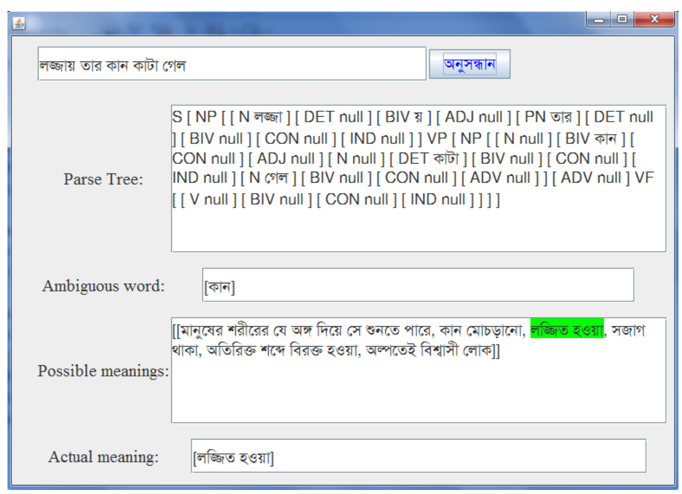


Figure 6: Snapshot of our system for the sentence "লজ্জায় তার কান কাটা গেল"

Input sentence: "লজ্জায় তার কান কাটা গেল"
Ambiguous word: কান
Actual meaning: লজ্জিত হওয়া

Now the snapshot for the output where the ambiguous word is used as an adjective is given below:

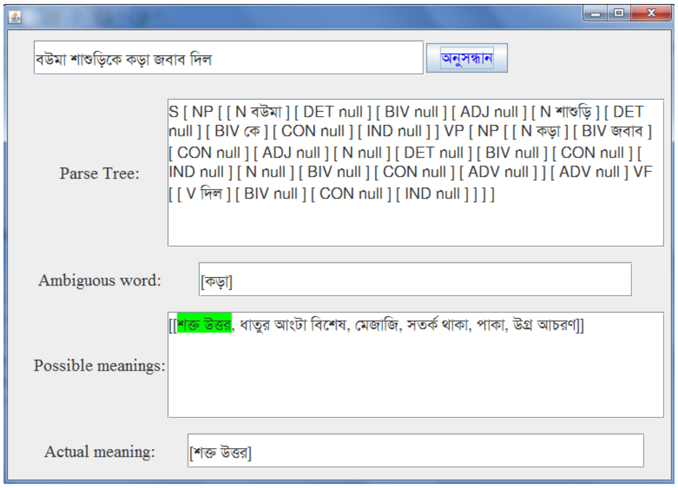


Figure 7: Snapshot of our system for the sentence "বউমা শাশুড়িকে কড়া জবাব দিল"

Input sentence: "বউমা শাশুড়িকে কড়া জবাব দিল"
Ambiguous word: কড়া
Actual meaning: শক্ত উত্তর

When we consider the ambiguous word as a verb then the snapshot of the resultant output as follows:

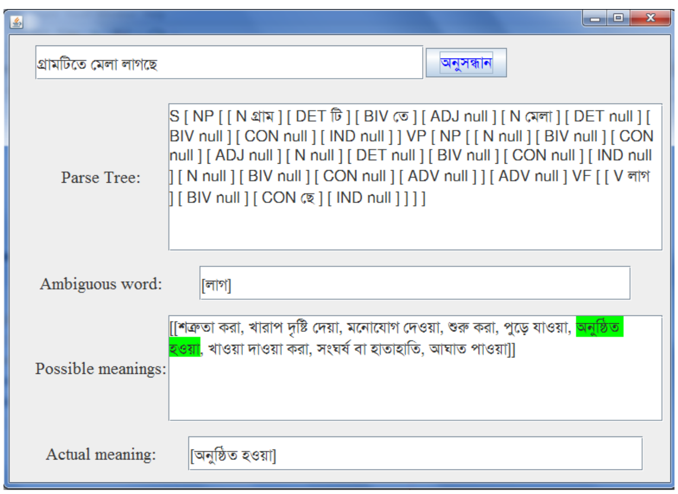


Figure 8: Snapshot of our system for the sentence "গ্রামটিতে মেলা লাগছে"

V. EXPERIMENTAL RESULTS

To evaluate the effectiveness and the overall performance rate of our proposed system, we have tested some sentences and determine the success rate of the system. We have selected our input sentences from Bangla grammar books [14,15] and also from the sources of Bangla text books, Internet, Bangla Blog Sites. After collecting the input Bangla sentences with ambiguous word we have tested our system for about 500 different sentences whether they are correctly performed or not.

Here we have tested our proposed system through quantitative measures and expressed in terms of percentage. The quantitative measurement is the measurement of data that can be put into numbers. The goal of quantitative measurement is to run statistical analysis, so data has to be in numerical form.

A. Success Rate

For the performance analysis, we have used the equation of success rate. Success rate denotes the ratio between total no. of correctly performed disambiguated sentences and total no of input sentences.

We have calculated the success rate for our system on the basis of input sentence length, types of ambiguous word and the number of the ambiguous word per sentence. The success rate for input sentence length is 86.97%, types of ambiguous word are 86.15% and for number of ambiguous word per sentence is 87.32%. As we have selected 500 input sentences and among them 412 sentences are generated the correct output so the overall performance of our proposed system is 82.40%.

Table 4, 5 and 6 illustrate the success rate sequentially for the input sentence length, different types of ambiguous word and the number of ambiguous word per sentence and their corresponding graphs are in figure 9,10 and 11.

TABLE IV. SUCCESS RATE ON THE BASIS OF VARIOUS LENGTH OF INPUT SENTENCES

Input sentence length	No of input sentences per length	Correctly performed sentence per length	Success Rate(%) per sentence length	Success Rate (%)
3	25	22	88	
4	50	45	90	
5	60	53	88.33	86.97
6	50	43	88	
7	30	24	80	

The corresponding graphical figure for this data table is given below:

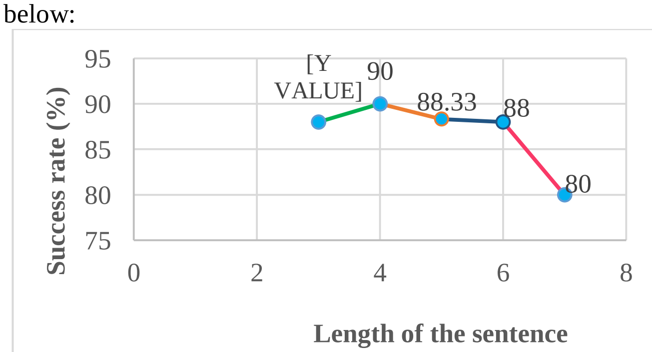


Figure 9: Success rate vs. input sentence length graph

TABLE 5. SUCCESS RATE ON THE BASIS OF THE TYPES OF AMBIGUOUS WORD

Types of ambiguous word	No of sentences per types	Correctly performed sentence per types of ambiguous word	Success Rate(%) per types of ambiguous word	Success Rate (%)
Noun	150	130	86.67	
Adjective	70	60	85.71	86.15
Verb	40	34	85	

The graphical representation of the success rate versus ambiguous word types is given below:

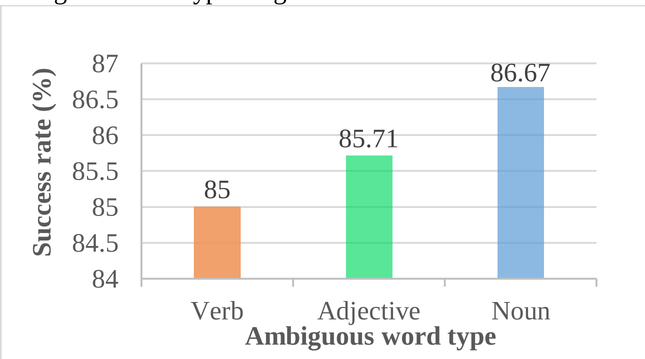


Figure 10: Success rate vs. ambiguous word type graph

Data table on the consideration of number of ambiguous word per input sentence as shown in Table VI.

TABLE VI. SUCCESS RATE ON THE BASIS OF NUMBER OF AMBIGUOUS WORD

Number of ambiguous word per sentence	No of sentences per types	Correctly performed sentences	Success Rate(%) per number of ambiguous word	Success Rate (%)
1	125	108	86.40	87.32
2	60	50	83.33	
3	20	16	80	

According the data table we get the graph as in Fig. 11.

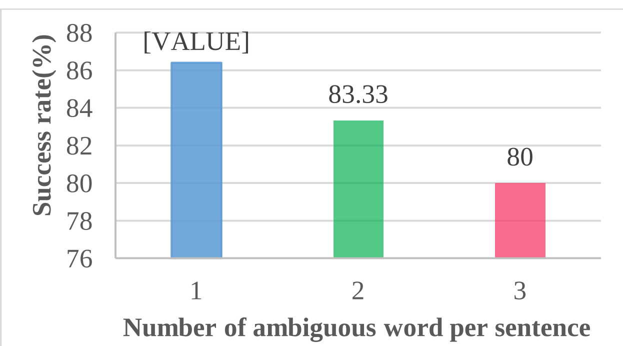


Figure 11: Success rate vs. number of ambiguous word graph

VI. Conclusion

Word sense disambiguation technique is one of the most important task in natural language processing. This task highly depends on lexical and syntactic information along with semantic information. The main objective of our work is to design a Bangla WSD System which can detect the ambiguous Bangla word and its actual meaning in a Bangla input sentence.

Hence, a good parser will play the major role at syntax level during disambiguation process and a parse tree can be generated after tokenization. As we have used a knowledge based approach to design our system, we need to build a machine readable dictionary for the verification of the tokens. After this verification we can construct a parse tree and check it for the detection of ambiguous word. With the help of this MRD we can easily find or detect the ambiguous word which is most overlapped with the dictionary definitions and also its actual meaning. Although the proposed Bangla WSD system is evaluated for limited sentences, the overall accuracy is 82.40% and it can only perform up to syntactic analysis. We have not considered the semantic features of a sentence. Current system can be extended by adding semantic features of the input sentences and also enrich the word database. These are left for future issues.

VII. References

[1] G. A. Miller, R. Beckwith, C. D. Fellbaum, D. Gross, K. Miller. 1990. WordNet: An online lexical database. Int. J. Lexicograph. 3, 4, pp. 235–244.

[2] N. N. Islam, N. Nawaz, M. Tanzil and Dr. M. M. Ali, "Supervised Information Theoretic Bangla Word Sense Disambiguation Using Lexeme List", Conference on Language and Technology (CLT07), August 2007, University of Peshawar, Pakistan.

[3] R. Pal, D. Saha and N. S. Dash, "Automatic Classification Of Bangla Sentences Based on Sense Definitions Present In Bengali WordNet",IJCTCM, Vol.5, No.1, January 2015.

[4] R. Pal, D. Saha, S. Naskar and N. S. Dash, "Word sense disambiguation in Bengali: A lemmatized system increases the accuracy of the result", Recent Trends in Information Systems (ReTIS), IEEE 2nd International Conference on9-11 July 2015.

[5] R. Pandit and S. K. Naskar, "A memory based approach to word sense disambiguation in Bengali using k-NN method",Recent Trends in Information Systems (ReTIS), IEEE 2nd International Conference on9-11 July 2015.

[6] D. McCarthy (2011), "Word Sense Disambiguation", University Of Melbourne.

[7] D. Yarowsky, "Unsupervised Word Sense Disambiguation Rivaling Supervised Methods," Proceedings of the 33rd annual meeting on Association for Computational Linguistics, pp.189-196, Cambridge, Massachusetts, 1995.

[8] R. Navigli, M. Lapata (2010), "An Experimental Study of graph connectivity for unsupervised word sense disambiguation", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 32, No. 4.

[9] S. Singh, V. K. Singh and T. J. Siddiqui, "Hindi Word Sense Disambiguation Using Semantic Relatedness Measure" Department of Electronics & Communication, University of Allahabad, Allahabad, India.

[10] R. Miller, G. A., "WordNet: An On-line Lexical Database," Communications of the ACM, Vol.38 No. 11, 1995.

[11] E. Agirre and P. Edmonds (2006), "Word Sense Disambiguation: Algorithms and Applications", Springer, New York, NY.

[12] E. Palta and P. O. Damani (2006-2007), "Word Sense Disambiguation: First Stage Report", IITB.

[13] M. M. Hoque and M. M. Ali, "A Parsing Methodology for Bangla Natural Language Sentences", In Proc. ICCIT'03, Dhaka, Bangladesh, pp. 277-282, vol. 2, 2003.

[14] Humayun Azad, Bakyatattya(বাক্যতত্ত্ব), The University of Dhaka, Dhaka, 1995.

Bangladeshi Vehicle Digital License Plate Recognition for Metropolitan Cities Using Support Vector Machine

Md Azher Uddin
Computer Science & Engineering
International Islamic University Chittagong
Chittagong, Bangladesh
azher006@yahoo.com

Joolekha Bibi Joolee
Computer Science & Engineering
International Islamic University Chittagong
Chittagong, Bangladesh
julekhajulie@gmail.com

Shayhan Ameen Chowdhury
Computer Science & Engineering
International Islamic University Chittagong
Chittagong, Bangladesh
shayhan@yahoo.com

Abstract— **Bangladeshi vehicle digital license plate recognition system using support vector machine for metropolitan cities (i.e. Dhaka, Chittagong) is presented in this paper. The proposed system divided into three major parts- license plate detection, plate character segmentation and character recognition. Experiments have been done for this proposed framework. More than 1000 images taken from various scenes are used, including diverse angles, different lightening conditions and complex scenes. In the first phase, Sobel operator and histogram analysis is used to detect the license plate region. Then, connected component labeling and bounding box technique used to segment the characters of detected license plate region. After that, Gabor filter is applied on the segmented characters to acquire desired character features. Since feature vector obtained using Gabor filter is in a high dimension, to reduce the dimensionality a nonlinear dimensionality reduction technique that is Kernel PCA has been used. Finally, Support Vector Machine has been used for classification. The experimental results show that proposed method can correctly recognize the license plate characters.**

Keywords-vehicle digital license plate detection; morphology; Sobel operator; connected component labeling; Gabor filter; Kernel PCA; Support Vector Machine.

I. INTRODUCTION

License Plate Recognition (LPR) is a vital research area due to its applications. The task of recognizing specific object in an image is one of the hardest topics in the field of computer vision. For license plate character recognition we can employ existing closed-circuit TV or road-rule enforcement cameras. License plate recognition can be used in many application such as electronic toll collection on pay-per-use roads, parking or exit managing and traffic control and management. Vehicle Retro-Reflective license plate commonly known as Vehicle digital license plate, the most current and technically improve license plate ever in Bangladesh. Government of Bangladesh has a goal to turn the country digital within 2021. As a part of digitalization, BRTA (Bangladesh Road Transport Authority) has introduced the Retro-Reflective license plate generally known as digital license plate.

The vehicle license plate recognition task is quite challenging from vehicle images due to the view point changes, and the non-uniform outside illumination conditions during image acquisition. In this paper we have used a database, which contains complex images. In the database, images also taken in different background, illumination and distance between the vehicle and camera is also different. This paper proposes a framework for Bangladeshi digital license plate recognition. The framework composed of three processing steps: 1) license plate detection by morphological processing; 2) segmentation of plate characters by connected component labeling and bounding box and 3) recognition of each character by extracting the features from the character and then recognizing by SVM classifier.

In segmentation phase, our method focuses on a solution for image disturbance resulting from non-uniform lighting condition and several outdoor conditions such as shadow. Additionally, our approach has numerous inherent advantages over other feature extraction technique. Our proposed framework is almost independent of slant or skewness. In feature extraction phase, features are extracted from the characters by Gabor filter, which is invariant to non-uniform lighting condition, rotation, scaling, and translation. Then Kernel PCA is used to reduce the high dimensionality of feature vector that is obtained by Gabor filter. Finally, SVM classifier is used to recognize the characters. The experiment results demonstrate a maximum recognition rate of 99.2%. Moreover, our experimental result also compared with previous works conducted for the recognition of Bangladeshi license plate.

The paper is organized as follows. The next section composes a review of related researches that have been implemented. Proposed framework is described in section III. Section IV presents results evaluation. Finally, conclusion is given in section V.

II. RELATED WORK

Different techniques are developed for license plate extraction. This section presents some previous related works for license plate detection which is significant to proposed approach.

Raiyan Abdul Baten et al [1] proposed a simple method for license plate recognition of Bangladesh using template matching. Here the authors briefly describe the characteristics of license plate in the metropolitan cities of Bangladesh. In [2], morphological operation used for license plate detection and segment modeling technique is used for recognizing the characters. In [3], HSI color model and geometrical properties used by the authors to detect the Bangladeshi vehicle license plate. In [4], authors proposed Bangla automatic number plate recognition system using artificial neural network. A robust feature extraction technique is applied to extract the feature from each characters which is invariant to the rotation and scaling.

In [5], a novel adaptive image segmentation method named as sliding concentric windows used for detecting license plate region. Kaushik Deb et al [6] proposed an efficient method using sliding concentric windows and artificial neural network to recognize the license plate. In [7], morphological processing and support vector machines used to recognize the license plate characters with an average accuracy of 97.89%. Morphology and least squares support vector machines used to recognize the license plate in [8]. In [9], edge detection and morphological processing used for license plate detection. Support vector machine used to recognize Chinese license plate in [10] with an average accuracy of 96.3%.

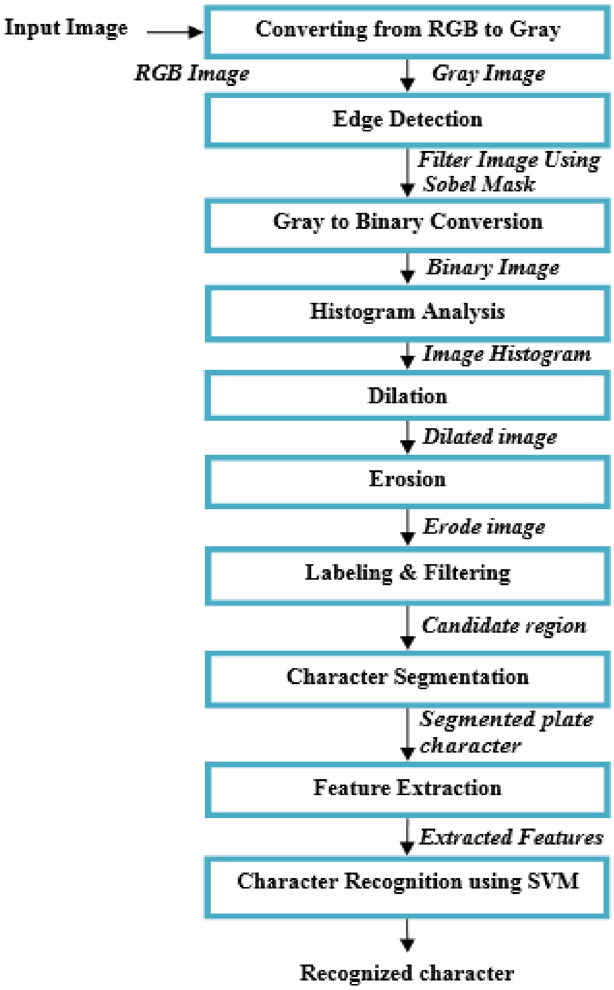


Fig. 1. Proposed Framework

III. PROPOSED FRAMEWORK

The proposed vehicle license plate recognition algorithm consists of three main stages: (1) detecting candidate region (2) character segmentation and (3) character recognition. The block diagram in figure 1 shows the working approach towards the solution of the stated problem.

A. RGB to Gray scale conversion

The vehicle with digital license plate is first captured using digital camera and stored in .jpg format. More than 1000 images taken from various scenes were used, including complex scenes, diverse angles and different lightening conditions. The size of the input images is 640×480 pixels. Figure 2 shows a capture image of vehicle with digital license plate.

The RGB color model consists of the three additive colors: red, green, and blue. This process converts the RGB image into gray scale image. We require to do this conversion because we will perform edge detection operation on gray scale image. The resulting gray scale image will be in two dimensional. The range will be among 0 to 255 values. The value 255 represents pure white and 0 represents pure black. Figure 3 shows RGB to gray conversion result. Conversion is done using weighted sum method as

$$GI = 0.2990*R + 0.5870*G + 0.1140*B \qquad (1)$$

Fig. 2. Captured RGB image

Fig. 3. RGB to gray scale image conversion result

B. Edge Detection

An edge represents the border line of an object which can be used to detect the shapes and area of the particular object. A license plate region contains many vertical edges due to characters and numbers. Many methods have been proposed for edge detection [2, 9-10]. We selected the sobel operator for finding the license plate region in an image because it produced better results and it also works faster compared to other gradient operators. Sobel filter creates an image which emphasizes edges and transitions. There are two masks for sobel filter, one is horizontal mask and another one is vertical mask. Vertical mask will find the edges in vertical direction, it is because the zeros column in the vertical direction. Horizontal mask will find edges in horizontal direction, as zeros column is in horizontal direction. Here we used sobel vertical mask as we need vertical edges. The result of edge detection is shown in figure 5.

-1	0	1
-2	0	2
-1	0	1

-1	-2	-1
0	0	0
1	2	1

Fig. 4. (a) Sobel vertical mask and (b) Sobel horizontal mask

Fig. 5. Vertical edge detection

C. Gray scale image to Binary image conversion

Then the gray scale image is converted to a binary image. A binary image is a digital image that has only two possible colors (i.e. black and white) for each pixel. Here we used Otsu method [11] to convert the processed image to binary image.

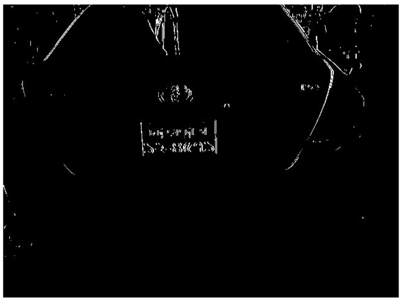


Fig. 6. Binary image

D. Histogram Analysis

Vertical projection refers to find out the sum of pixels in each column of an image. And horizontal projection refers to find out the sum of pixels in each row of an image. Here, histogram analysis had done to find the vertical projection of the processed binary image. In vertical projection, y-axis is the rows of the image, and x-axis shows the number of white pixels in each row. Figure 7 shows the resulted vertical projection of the binary image in figure 6. In the vertical projection, the rows parallel to the license plate region generally have the maximum values. Then based on a threshold value t, we find out the rows with the maximum values in the vertical projection.

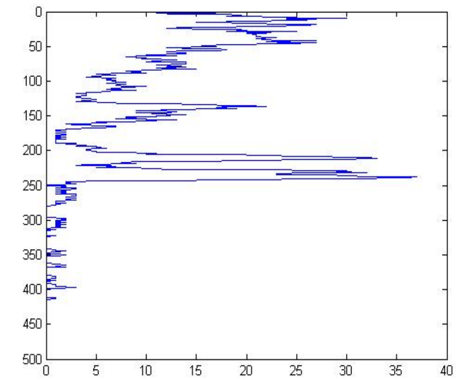


Fig. 7. Vertical projection of the binary image

E. Dilation

Morphological operations apply a structuring element to an input image, producing an output image of the same size. The most fundamental morphological operations are dilation and erosion. Dilation inserts pixels to the borders of objects in an image, while erosion removes pixels on object borders.

The dilation of object A by structuring elements B can be defined by:

$$A \oplus B = \bigcup_{b \in B} A_b \tag{2}$$

Here we first dilated the processed image horizontally and then vertically. From these two dilated images we find out the common white pixels. Again dilation operation is applied on the processed binary image. The structuring element for all dilations are rectangles. The probable holes are filled as we want to get a continuous area for license plate region. The result of dilation is shown in figure 8.

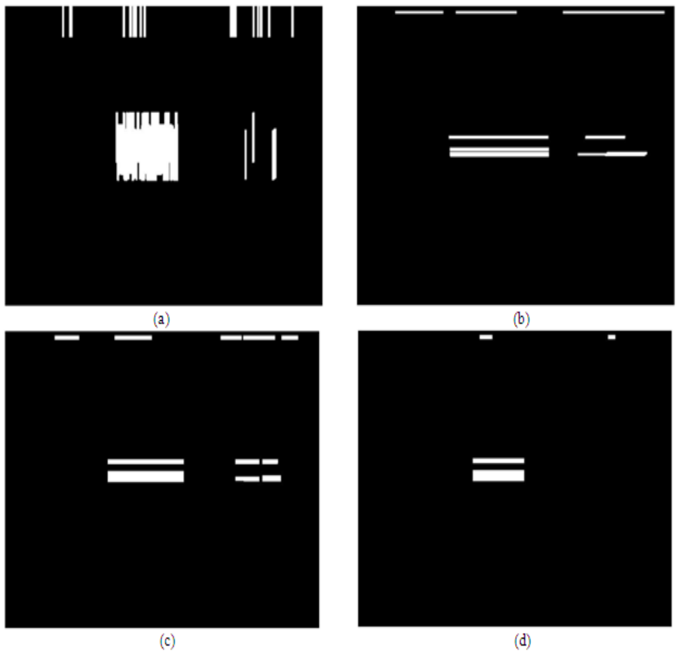


Fig. 8. (a) Image after Horizontal dilation, (b) Image after Vertical dilation (c) Image after dilation process (d) Image after Erosion process

F. Erosion

Erosion operator excludes any extra regions, which do not belong to the plate. The erosion of the binary image A by the structuring element B can be defined by:

$$A \ \Theta \ \mathrm{B} = \bigcap_{b \in B} A_{-b} \tag{3}$$

Here we erode the processed binary image with a horizontal line. The result of erosion is shown in figure 8(d).

G. Labeling and Filtering

Connected-component labeling is used in computer vision to identify connected regions in binary digital images. Next step is to find biggest binary region. For this first we find out the number of connected regions using connected component labeling [12]. Then we compute the area of every regions and find out biggest binary region i.e. license Plate. Based on a threshold value we enlarge the candidate region due to Bangladeshi license plate contains two lines of characters and digits. Figure 9 shows the detected license plate region. Then license plate region is extracted from the image by using image crop function from the MATLAB.

Fig. 9. Detected license plate region

H. Character Segmentation

Character segmentation is a process that seeks to decompose an image of a sequence of characters into sub-images of individual symbols. Here the extracted candidate region image is a RGB image. After extracting the candidate region, image is converted to binary image using adaptive thresholding.

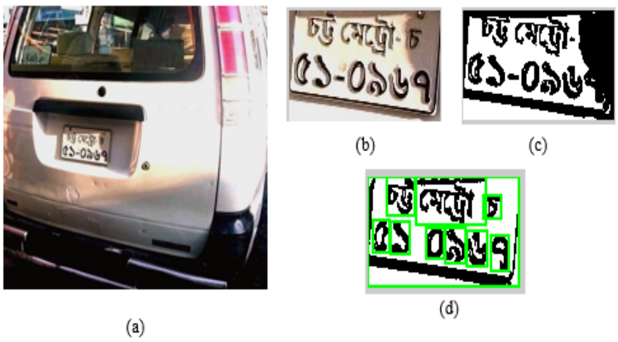


Fig. 10. (a) Input image; (b) Extracted license plate; (c) Binary conversion by traditional method; (d) Binary conversion by adaptive thresholding

Adaptive thresholding is used to overcome non-uniform illumination problem. Due several outdoor conditions, shadow may appear in the license plate region during image acquisition, which are usually challenging for obtaining successful processed results using traditional binary techniques. The result of adaptive thresholding is compared with traditional binary conversion method (i.e. Otsu method) in figure 10. The whole process of adaptive thresholding algorithm consists of several steps:

1. Convolution operation is performed on the image by using a suitable statistical operator, i.e. mean.
2. Deduct the original image from the convolved image.
3. Threshold the difference image with C (i.e. C is a constant).

Once the candidate area is binarized the next step is to segment the characters. At first, Remove all object containing fewer than 30 pixels to eliminate regions without interest such as small noisy regions. Then we label connected components in candidate region. After that we measure a set of geometrical properties such as bounding box and aspect ratio for each connected components (object) in the binary image. Finally, characters are extracted from the candidate region by aspect ratio. The aspect ratio [3] can be found by the ratio of width and height of the bounding box of an object. The aspect ratio is defined by,

$$AR = \frac{(C_{max} - C_{min}) + 1}{(R_{max} - R_{min}) + 1} \tag{4}$$

Where C and R indicate columns and row, respectively. Objects which satisfy ***AR*** (aspect ratio) bounds 1 to 2 are considered as candidate characters. The result of character segmentation is shown in figure 12.

Fig. 11. (a) candidate region and (b) after labeling connected components candidate region

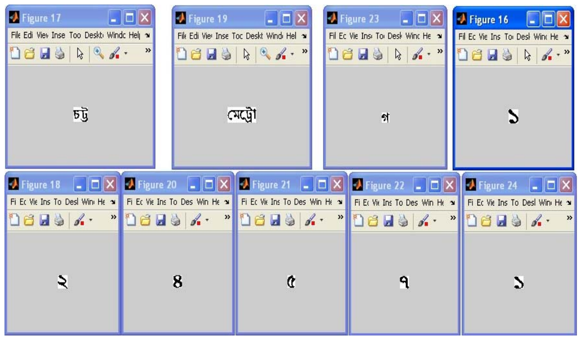


Fig. 12. Character segmentation

I. Feature Extraction

Selecting good features is a critical phase in any object recognition system. Here the segmented binary image characters are resized to 50×50. In this phase, we employed a Gabor filter in order to obtain a feature vector. Gabor filtering is successfully used in many image processing and analysis domains such as: image smoothing, shape analysis, texture analysis, face recognition, fingerprint recognition and iris recognition. The Gabor filter is a band-pass filter whose impulse response is defined by a harmonic function multiplied by a Gaussian function. Thus, a bi-dimensional Gabor filter constitutes a complex sinusoidal plane of particular frequency and orientation modulated by a Gaussian envelope [13]. The most significant advantage of Gabor filters is their invariance to illumination, rotation, scaling, and translation. Here to overcome the problem of slant or skewness, Gabor filter is used. In the spatial domain, a two-dimensional Gabor filter [14], defined as:

$$G(x,y) = \frac{f^2}{\pi\gamma\eta} exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right) exp(j2\pi f x' + \phi) \quad (5)$$

$$x' = xcos\theta + ysin\theta \quad (6)$$

$$y' = -xsin\theta + ycos\theta \quad (7)$$

Where, f is the frequency of the sinusoidal factor, σ is the standard deviation of the Gaussian envelope, ϕ is the phase offset, θ represents the orientation of the normal to the parallel stripes of a Gabor function and γ is the spatial aspect ratio which specifies the ellipticity of the support of the Gabor function. Our proposed system employs thirty-two Gabor filters in four scales and eight orientations. Then downsampling is done, here column downsampling factors is (5×5). Thus the size of the Gabor feature vector obtained is (50×50×4×8) / (5×5) which is 3200. Even after downsampling the feature vector obtain using Gabor filter is still large. Therefore, dimensionality reduction methods have been employed. For dimensionality reduction Kernel PCA (KPCA) has been employed. KPCA is a nonlinear dimensionality technique. KPCA is an extension of PCA. Principle component analysis (PCA) is a conventional linear feature extraction method mostly used in numerous pattern recognition method, for example [15]. Kernel oriented approaches have been used in various applications for example in [16-17]. KPCA maps the features from a low dimensional space to a more high dimensional space. By doing this, the features become linearly separable**.** Kernel PCA [18] calculates the principal eigenvectors of the kernel matrix, instead of those of the covariance matrix. The reconstruction of conventional PCA in kernel space is easy, due to a kernel matrix is similar to the inproduct of the features in the high-dimensional space that is constructed using the kernel function. The application of PCA in kernel space gives Kernel PCA the property of constructing nonlinear mappings. Kernel PCA calculates the kernel matrix K of the features x_i. The kernel matrix can be defined by,

$$k_{ij} = \kappa(x_i, x_j) \quad (8)$$

where κ is a kernel function. Then, the kernel matrix K is centered using the following alteration of the entries

$$k_{ij} = k_{ij} - \frac{1}{n}\sum_l k_{il} - \frac{1}{n}\sum_l k_{jl} + \frac{1}{n^2}\sum_{lm} k_{lm} \quad (9)$$

The centering process corresponds to deducting the mean of the features in conventional PCA. It makes guaranteed that the features in the high-dimensional space defined by the kernel function are zero-mean. Then, the principal d eigenvectors v_i of the centered kernel matrix are calculated. It can be shown that, the eigenvectors of the covariance matrix α_i are scaled forms of the eigenvectors of the kernel matrix v_i

$$\alpha_i = \frac{1}{\sqrt{\lambda_i}} v_i \quad (10)$$

In order to get the low-dimensional data representation, the feature is projected onto the eigenvectors of the covariance matrix. The outcome of the projection (i.e., the low-dimensional data representation Y) is given by,

$$Y = \left\{\sum_j \alpha_1 \kappa(x_j, x), \sum_j \alpha_2 \kappa(x_j, x), \ldots, \sum_j \alpha_d \kappa(x_j, x)\right\} \quad (11)$$

where κ is the kernel function, used in the calculation of the kernel matrix. Since Kernel PCA is a kernel-based technique, the mapping done by Kernel PCA vastly depend on the choice of the kernel function κ. Probable selections for the kernel function include the linear kernel, the polynomial kernel, and the Gaussian kernel. In our work, we used Gaussian kernel function. The dimension of the feature vector has been reduced to 140 from 3200 using Kernel PCA.

J. Character Recognition

The training and classification processes are done by using Support Vector Machine (SVM) classifier. SVM [19] is a classification prediction tool that uses machine learning concept to maximize predictive correctness while spontaneously avoiding over-fit to the data.

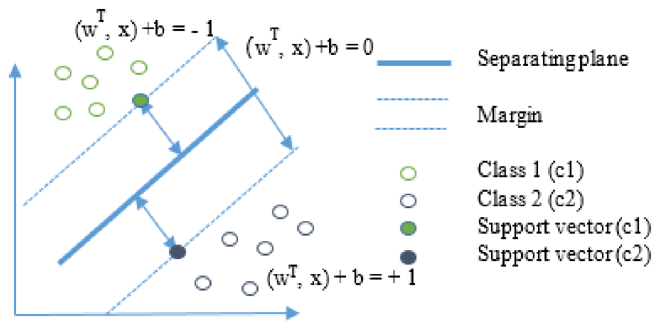


Fig. 13. Linear separating hyperplanes for the separable case

We are given a set of n data points $\{x_i, y_i\}$; where, $i = 1,\ldots,$ n, $y_i \in \{-1, +1\}$ and $x_i \in R^n$. SVM method targets at finding a classifier of form,

$$y(x) = sign\left(\sum_{i=1}^{n} \alpha_i y_i k(x, x_i) + b\right) \quad (12)$$

Where α_i are positive real constants and b is a real constant; $k(x, x_i)$ is the kernel function that can be defined by,

$$k(x, x_i) = \varphi(x)^T \varphi(x_i) \quad (13)$$

φ is a nonlinear mapping function used to map input data point x_i into a higher dimensional space. In our proposed framework, the Gaussian radial basis function [7] is used. It is defined by,

$$k(x, z) = e^{-\gamma \|x - z\|^2} \quad (14)$$

In the high dimensional space we consider that, the data can be separated by a linear hyperplane, according to the following equations,

$$\begin{cases} w^T.\ x_i + b \geq +1, & if\ y_i = +1 \\ w^T.\ x_i + b \leq -1, & if\ y_i = -1 \end{cases} \quad (15)$$

In our work, we build two big SVM classifiers for numbers and alphabets respectively. Both classifier implements binary tree structure and applies one against rest to build sub-classifiers. Let us consider numeral classifier as an instance, it contains 9 sub-classifiers. Figure 14 shows the numeral classifier. Each layer contains a digit as a leaf node and a sub-classifier except for last layer that contains two digits as a leaf nodes and the root node is a classifier. Each leaf node is a digit and a positive output ($y_i = +1$), on the other hand each sub-classifier is a negative output ($y_i = -1$).

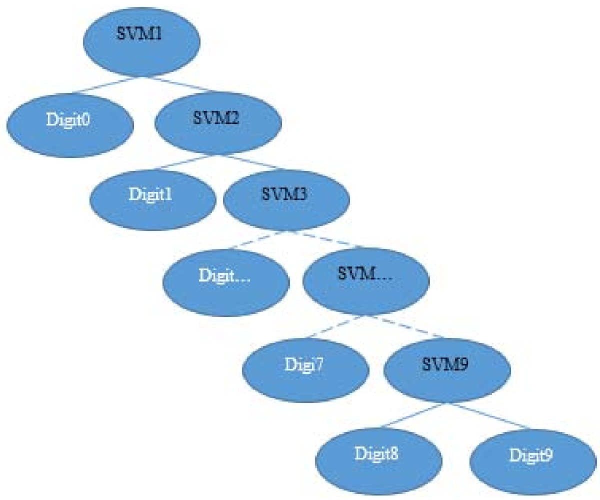


Fig. 14. Binary tree based numeral classifier

IV. RESULT EVALUATION

All experiments were done on dual-core 3.00 GHz with 2 GB RAM under MATLAB environment. In the experiments, we have used more than 1000 images with the size 640 × 480 pixels. The images were taken from different illuminations, diverse angles and complex scenes. Some example images are shown in Fig. 15. The images for training and testing are unrelated. The database is separated into two datasets. The first dataset contains 100 images with total of 900 characters and are used for training the SVM character classifiers. On the other hand, another dataset contains 917 images with total of 8253 characters and are used for testing the performance of the classifiers. Nevertheless, for license plate detection the entire database was used. The license plate detecting rate of success is 93.2%. And the license plate character recognition rate of success is 99.2%. Results of license plate detection for proposed work in different conditions are shown in table I.

Fig. 15. Example images in different illumination, complex scenes and diverse angles

TABLE I. DETECTION RESULT IN DIFFERENT CONDITION

Condition	No. of Images	Ex. LPs	Success rate (%)
Different illuminations	305	292	95.7 %
Diverse angles	206	195	94.6 %
Complex scenes	106	98	92.4 %
Various Environments	400	363	90.8 %
Total	1017	948	93.21 %

Recognition rate (RR) is calculated as

$$\text{RR} = \left(\frac{\text{No. of recognized samples}}{\text{No. of total samples of that sign}}\right) \times 100\% \quad (16)$$

For example, recognition rate of Bangla number '১' (ek) = (91/91)*100 = 100%

Recognition rate for all Bangla characters =

(8187/8253)*100 = 99.2%

In [1], the system can work very satisfactorily for samples that are not too noisy and not over skewed, which is not practical. Our proposed approach solve the problems of [1].

Table II shows the comparison among applied method and other well reported methods for license plate detection, character segmentation and character recognition. The total process success rate for our proposed method is 91.3%. Figure 16 shows the comparison of overall system performance between our proposed method and other related works. So from the table II and figure 16 we can see that, our proposed method outperforms the existing methods for Bangladeshi license plate recognition.

TABLE II. COMPARISON AMONG APPLIED METHOD AND OTHER WELL REPORTED METHODS FOR LICENSE PLATE DETECTION, CHARACTER SEGMENTATION AND CHARACTER RECOGNITION

Reference	License Plate Detection Accuracy	Character Segmentation Accuracy	Character Recognition Accuracy
Proposed Method	93.2 %	98.1 %	99.2 %
[2]	88.0 %	98.0 %	98.0 %
[3]	84.8 %	-	-
[4]	92.1 %	97.5 %	84.2 %
[7]	97.6 %	90.7 %	97.9 %

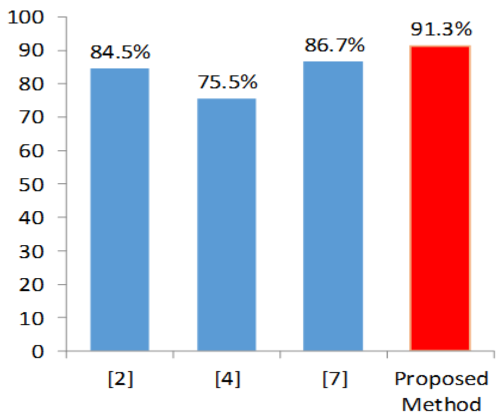


Fig. 16. Comparison of overall system performance between our proposed method and other related works

V. CONCLUSION

This paper presents a vehicle license plate recognition method based on support vector machine. Here, first Sobel operator and histogram analysis is used to detect the license plate region. Subsequently, connected component labeling and bounding box method used to segment the characters. Finally, features are extracted from the segmented character and then character is recognized using support vector machine. During the experiment, different illumination conditions, diverse angles and varied distances between vehicle and camera often occurred. In such cases, the result is very effective when the proposed method is used.

REFERENCES

[1] Raiyan Abdul Baten, Zunaid Omair and Urmita Sikder, "Bangla License Plate Reader for Metropolitan Cities of Bangladesh Using Template Matching", 2014 International Conference on Electrical and Computer Engineering (ICECE), pp.776-779, 2014.

[2] Nahian Alam Siddique, Asif Iqbal, Fahim Mahmud and Md. Saifur Rahman, "Development of an Automatic Vehicle License Plate Detection and Recognition System for Bangladesh", 2012 International Conference on Informatics, Electronics & Vision (ICIEV), pp. 688-693, 2012.

[3] Kaushik Deb, Muhammad Kamal Hossen, Muhammad Ibrahim Khan, Mohammad Rafiqul Alam, "Bangladeshi Vehicle License Plate Detection method based on HSI color model and geometrical properties", International Forum on Strategic Technology (IFOST), PP. 1-5, 2012.

[4] Md. Mahmudul Alam Joarder, Khaled Mahmud, Tasnuva Ahmed, Mohsina Kawser and Bulbul Ahmed,"Bangla Automatic Number Plate Recognition Using Artificial Neural Network", Asian Transactions on Science and technology, vol. 02, Mar. 2012.

[5] Kaushik Deb, Hyun-Uk Chae and Kang-Hyun Jo, "Vehicle License Plate Detection Method Based on Sliding Concentric Windows and Histogram", Journal of Computers, VOL. 4, NO. 8, pp. 771-777, 2009.

[6] Kaushik Deb, Md. Ibrahim Khan,Anik Saha and Kang-Hyun Jo, "An Efficient Method of Vehicle License Plate Recognition Based on Sliding Concentric Windows and Artificial Neural Network", Procedia Technology, Elsevier, pp. 812–819, Volume 4, 2012.

[7] Ayman Rabee and Imad Barhumi, "License Plate Detection and Recognition in Complex Scenes Using Mathematical Morphology and Support Vector Machines", 2014 International Conference on Systems, Signals and Image Processing (IWSSIP), pp. 59-62, 2014.

[8] Haiyan Zhao, Chuyi Song, Haili Zhao and Shizheng Zhang, "License Plate Recognition System Based on Morphology and LS-SVM", IEEE International Conference on Granular Computing, 2008, pp. 826-829, 2008.

[9] Aili Wang, Xusheng Liu, Yuhui Han and Changyan Qi, "License Plate Location Algorithm Based on Edge Detection and Morphology", 7th International Forum on Strategic Technology (IFOST), pp. 1-4, 2012.

[10] Xiaojun Chi, Junyu Dong , Aihua Liu and Huiyu Zhou, "A Simple Method for Chinese License Plate Recognition Based on Support Vector Machine", 2006 International Conference on Communications, Circuits and Systems Proceedings, pp. 2141-2145, 2006.

[11] N. Otsu, "A threshold selection method from gray-scale histogram". IEEE Transactions on System, Man, and Cybernetics, Vol. 9, pp. 62-66, 1979.

[12] Haralick, Robert M., and Linda G. Shapiro, Computer and Robot Vision, Volume I, Addison-Wesley, 1992, pp. 28-48.

[13] J.R. Movellan, "Tutorial on Gabor Filters," 2008. Available: http://mplab.ucsd.edu/tutorials/gabor.pdf [Accessed: December 12th 2014].

[14] M. Haghighat, S. Zonouz and M. Abdel-Mottaleb, "Identification using encrypted biometrics", Computer analysis of images and patterns (CAIP), Springer, pp. 440–448, 2013

[15] M. Kirby and L. Sirovich, "Application of the Karhunen-Loeve procedure for the characterization of human faces," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 12, pp. 103- 108, 1990.

[16] Milad Moghaddam, Manoochehr Nahvi and Reza Hassanzadeh Pak, "Static Persian Sign Language Recognition Using Kernel-Based Feature Extraction", 7th Iranian conference on Machine Vision and Image Processing (MVIP), pp. 1-5, 2011.

[17] A. Kocsor and L. Toth, "Kernel-based feature extraction with a speech technology application," Signal Processing, IEEE Transactions on, vol. 52, pp. 2250-2263, 2004.

[18] L.J.P. van der Maaten, E.O. Postma, and H.J. van den Herik, "Dimensionality Reduction: A Comparative Review. Tilburg University Technical Report", TiCC-TR 2009-005, 2009.

[19] Christopher J. C. Burges,"A tutorial on support vector machines for pattern recognition," Data Mining and Knowledge Discovery, Vol. 2, pp.121-167,1998.

Generalized Versions of Kirchhoff's Laws for Students

Shahriar Khan
Dept of EEE
Independent Univ., Bangladesh
khandhaka@gmail.com

Ziauddin Alamgir
Bangladesh Navy

Samina Alam
Premier University
Chittagong, Bangladesh

***Abstract-* Kirchhoff's Current Law is mostly taught for a single node, leaving uncertainties in students, sometimes in the simplest of circuits. It is proposed that KCL be shown as the current should equal zero across any cross-section of a circuit. KVL is mostly taught for a single loop. It is better shown early that the sum of voltages should equal zero around any possible loop in a multi-loop circuit. Students are mostly shown circuits that are "possible," with little mention of "impossible" circuits. However, students may unknowingly implement an "impossible" short circuit in the lab, or be unable to identify the impossibility of a given circuit. The proposed identification of impossibilities in circuits will provide students warning of short-circuits, and better insight into circuits in general. These approaches may be used also for AC and time-varying circuits. These practices yielded good results in the classroom in courses on DC circuits, AC circuits, and Systems and control.**

***Index Terms*-KCL, KVL, Circuits.**

I. INTRODUCTION

A good understanding of KVL and KCL is a prerequisite for studying DC and AC circuits. KVL and KCL continue in the 3rd or 4th year study of systems, such as non-periodic currents $i(t)$ and voltages $v(t)$. KVL and KCL have been widely explored by educators and researchers [1 - 4]. The original implications of KVL and KCL have been extended to numerous other fields [5 - 7].

The frequently-taught single-node version of KCL leaves the student with confusion and vagueness. Also KVL is shown mostly for a single loop, which leaves the student inadequately prepared for multiple node circuits and multiple loop circuits. To overcome the vagueness arising from present methods of teaching KCL and KVL, some modifications have been suggested [8 - 10]. These approaches to Kirchhoff's laws have been rarely documented, if at all, in the past.

The majority of books show only "possible" circuits, meaning they do not violate KVL and KCL. A student may easily implement an "impossible" circuit, or a short circuit in the lab, in violation of KVL. Or else, the student may not be able to identify the violation of KCL in a circuit. Asking students to identify the impossibility in given circuits is helpful in building insight and understanding. The material is largely based on the author's experience in the teaching of related subjects over a number of years. Some of the material was also published and followed in at least three books on the subjects [8,9,10]

II. SINGLE-NODE VERSION OF KCL

Following from the Law of conservation of charge, Kirchhoff's Current Law is most commonly stated that the sum of the currents flowing into a node is zero.

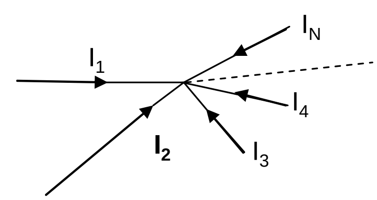


Figure 1. Students are taught the single-node version of KCL the sum of the currents flowing into a node equals zero.

This is the practice followed in the majority of text-books, at least for undergraduates. Most problems with KCL are taught in DC circuits. Less common is KCL with AC circuits, such as a problem as below.

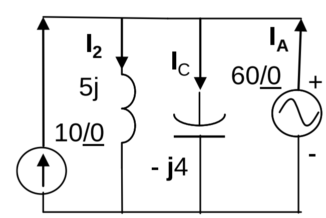


Figure 2. Single-node KCL for an AC circuit, solvable without a calculator

Simple application of KCL, without a calculator, shows that $\mathbf{I}_2 = -12\mathbf{j}$, $\mathbf{I}_C = 15\mathbf{j}$, and $\mathbf{I}_A = 3\mathbf{j} - 10$. The problem may be solved in quiz environment, and the current source, although rare in practice, helps develop concepts in the student.

A. *Problems from Single-node KCL*

As a result of the limited scope of the conventional KCL, students begin encountering vagueness in some of the simplest circuits they encounter.

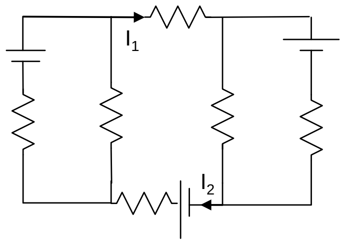
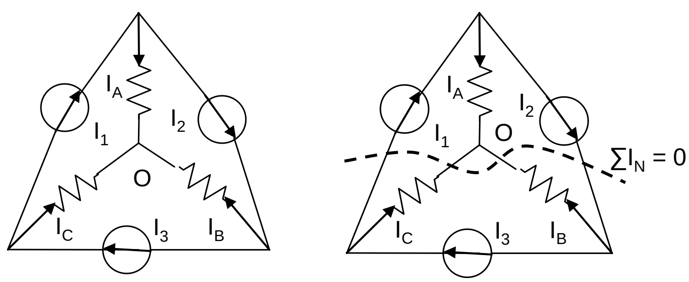


Figure 3. Students taught the single-node version of KCL are unclear about how $I_1 = I_2$.

In the circuit above, the single-node version of KCL does not directly show how I_1 and I_2 are equal. One reason is that there are at least two nodes (and not a single node) that is involved. The proposed cross-sectional version of KCL, as proposed in this paper, easily shows the beginning undergraduate that $I_1 = I_2$.



(a) **(b)**

Figure 4. KCL is not violated at node O in (a), regardless of the values of I_1, I_2, and I_3. The cross-sectional version of KCL (b) shows it more easily than the single-node version of KCL at node O.

In Figure 3(a) above, the single-node KCL does not make it readily apparent that KCL is not violated at node O for all values of I_1, I_2 and I_3. The validity has to be verified through the extra process of applying single-node KCL as follows:

$$I_A + I_B + I_C = I_1 - I_2 + (I_2 - I_3) + (I_3 - I_4) = 0$$

However, if KCL is applied across the cross-section shown, the circuit is seen to be possible at all times.

$$I_1 - I_2 + I_2 - I_3 + I_3 - I_1 = 0$$

Thus the cross-sectional version of KCL gives a shorter and clearer proof that KCL is not violated at the central node of the given circuit.

III. Proposed Cross-sectional Version of KCL

As a solution to the problems encountered above, it is proposed that KCL be additionally taught that,

"*The sum of the currents flowing through any cross-section of a circuit equals zero*."

This version of KCL follows also follows naturally from the Law of Conservation of Charge. More specifically, the charges per second crossing right across the cross section will equal the sum of the charges crossing left across the cross section.

A. *Applications of Cross-sectional KCL*

As a result of this new version of KCL, we can easily analyze the following circuits.

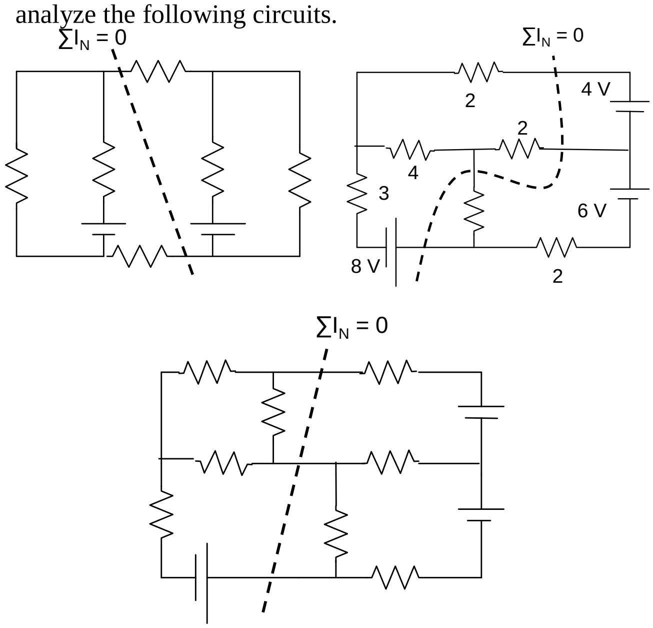


Figure 5. The circuits above may be analyzed by the proposed cross-sectional KCL

In the following circuits the single-node version does not readily allow us to find the unknowns V_1 and V_2.

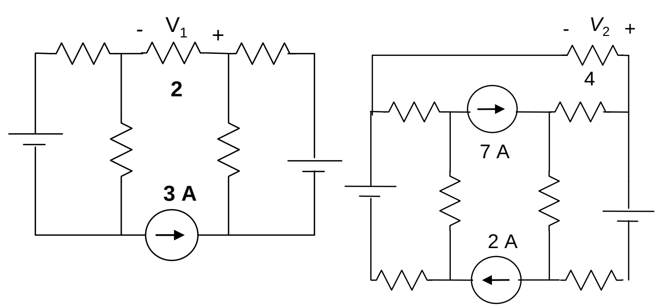


Figure 6. What is V_1 and V_2? The cross-sectional KCL readily shows that V_1 = 2x3 = 6 V, and V_2 = 4x5 V.

In the above circuits, using the proposed Cross-sectional KCL, V_1 is found.

$$V_1 = 2 \text{ x } 3 = 6 \text{ V}$$

Similarly, in (b), V_2 is found. $V_2 = 4(7 – 2) = 20$ V

The above concepts have been extended below to AC circuits, with current sources.

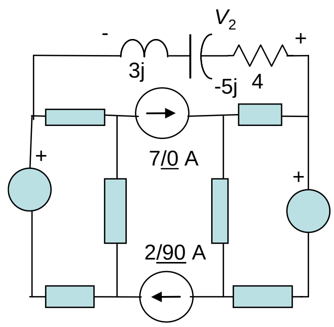


Figure 7. Problem in AC circuit that can be done

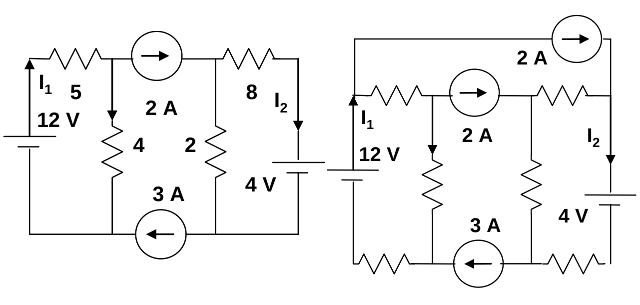
The voltage V_2 can be found using the cross-sectional version of KCL.

$$V_2 = (7 – 2\mathbf{j})(3j – 5\mathbf{j} +4) = (7 – 2\mathbf{j})(-2\mathbf{j} + 4) = 24 – 22\mathbf{j}$$

The current sources, although rare, and variety in the above circuits will give the undergraduate more insight, and better conception.

B. *Impossible Circuits*

Most books and curriculums focus on circuits that are actually possible. To introduce the student to a variety of concepts, it is also proposed that “impossible” circuits be included in books [8]. The single-node KCL does not illustrate the impossibility of the circuits in following figures. The new version of KCL readily shows the impossibility of the following circuits.



(a) (b)

Figure 8. Are these circuits possible? The proposed cross-sectional KCL illustrates the impossibility of the above circuits more easily than the conventional single-node KCL.

In fig. 8(a), from the cross-sectional version of KCL, it is apparent that the 2A and 3A current sources contradict each other, as the currents through the cross-section do not add up to zero.

In Figure 8(b) above, it is not readily apparent that something is wrong, from the conventional single-node version of KCL. From the updated cross-sectional version of KCL, it is apparent from the that the three sources 2A, 2A, and 3A contradict each other, as the currents through a central cross-section do not add up to zero. Thus the student gains insight into circuits, from the cross-sectional version of KCL.

IV. Single-Loop Version of KVL

Most text-books and curriculums teach Kirchhoff’s Voltage law that the sum of voltage rise around a single closed loop equals zero.

$$\sum E_N = \sum V_M$$

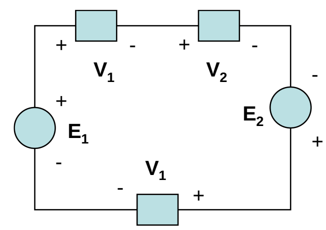


Figure 9. Most students are taught the above Kirchoff's Voltage Law for a single loop.

KVL follows from the law of conservation of energy,

A. *Problems Arising from Single-loop KVL*

The first problem that arises is that students are rarely told that KVL is applicable in each and every loop in a multiple-loop circuit.

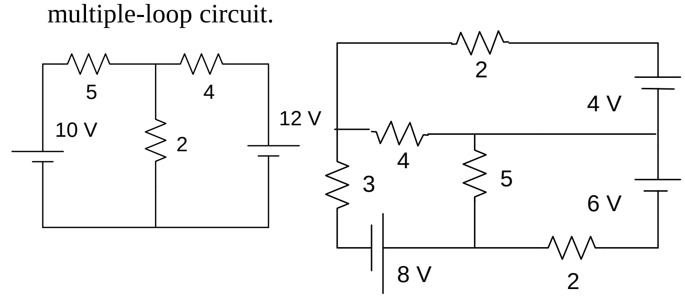
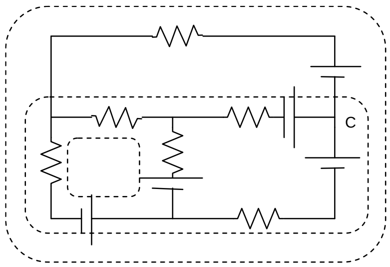


Figure 10. After learning the single-loop KVL, students are expected to understand that KVL is applicable around each and every possible loop in the above circuits.

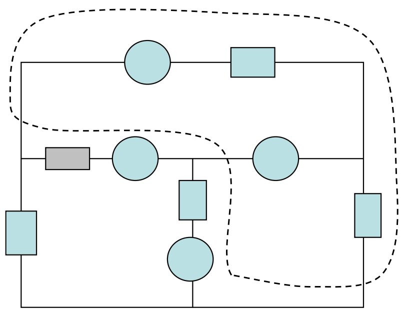
Immediately after being shown KVL for a single isolated loop, they are usually expected to understand that KVL is applicable for each and every loop in a multiple-loop circuit. For example, in the loop-current method, they are expected to apply the single-loop KVL to a multiple-loop circuit, without the clarification that KVL is applicable around every possible loop in a multiple loop circuit.

V. MULTIPLE-LOOP KVL

In most descriptions of KVL, students are not clearly told that in a circuit with multiple loops, KVL is satisfied in each and every loop. Three example loops have been shown in figure 9(a).



(a)



(b)

Figure 11. KVL is applicable around any loop of a network.

The only exceptions when KVL and KCL are not applicable are when there is a conflict, such as a short circuit in the network.

A. Applications

Using the multiple-loop version of KVL, we can easily find the currents below.

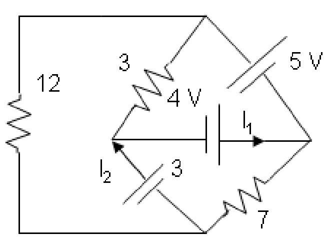


Figure 12. The multiple-loop version of KVL readily finds the currents I_1 and I_2.

To find: I_1= ___, I_2 = _____, we apply KVL around the various loops in the network. The current through the 3 ohm resistor equals,

$$(5 + 4)/3 = 9/3 = 3 \text{ A}$$

Current through 12 ohm resistor

$$(3 + 4 + 5)/12 = 12/12 = 1 \text{ A}$$

Current thorough 7 ohm resistor

$$(3 + 4)/7 = 7/7 = 1 \text{ A}$$

$$I_2 = 1 + 1 = \mathbf{2\ A}$$

$$I_1 = 3 + 2 = \mathbf{5\ A}$$

The circuit below is an AC version of the above circuit.

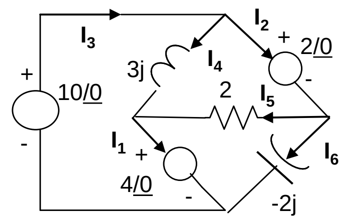


Figure 13. Multiple-loop AC circuit, that can solved without a calculator.

Applying KVL around the multiple-loop shows that,

$I_4 = (10 – 4)/3j = -2j$ A, $I_5 = (10 – 2 – 4)/2 = 2$

$I_6 = (10 – 2)/(-2j) = 4j$ A $I_1 = -2j + 2$,

$I_2 = 2 + 4j$, $I_3 = -2j + 2 + 4j = 2 + 2j$

In the circuit below, we find I_1, I_2 with the multiple-loop version of KVL.

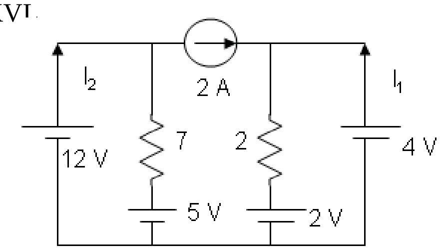


Figure 14. Applying KVL around the various loops allows us to calculate currents I_1 and I_2.

The multiple loop version allows us to calculate

$$I_2 = 1 + 2 = 3 \text{ A}$$

$$I_1 = 1 - 2 = -1 \text{ A}$$

Impossible Circuits

Most books and curriculums normally confine their attention to circuits that do not violate KVL. There is no mention of circuits that violate KVL and are therefore impossible. However, as lab instructors know, students are likely to implement short circuits in the lab, that are not mentioned in the books owing to their "impossibility," and violation of KVL. To broaden the insight of students, and to warn them about short-circuits, it is proposed here that students also be shown examples where KVL is violated.

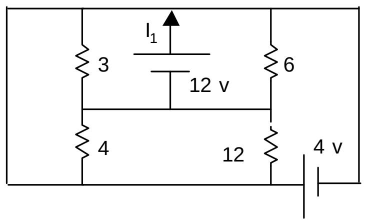


Figure 15. Application of the multi-loop version of KVL readily shows the short-circuits above, or violations of KVL.

With the circuits above, the student will become trained in identifying short-circuits, or circuits that violate KVL.

VI. Time-varying Circuits, and AC Circuits

In undergraduate curriculums in electrical engineering, KVL and KCL are shown mainly for DC circuits, to some extent for AC circuits, and rarely for other time-varying circuits (such as transients). Applications such as node-voltage, loop-current, and superposition methods, delta-wye transformations, and Thevenin's equivalent circuits are seldom shown for AC circuits, and almost never for other time-varying circuits. It is proposed here that cross-sectional KCL, multiple-loop KVL, and related applications be shown appropriately for AC circuits, and other time-varying circuits in undergraduate curriculums.

Conclusion

KVL and KCL, with accompanying node-voltage, loop-current, and the superposition method are important foundations in electrical engineering education.

The cross-sectional version of KCL resolves the vagueness, otherwise present in the single-node version of KCL. Problems with KCL for AC circuits have been shown, that are simple enough to be solved without a calculator. Similarly, the single loop version of KVL, popular in textbooks, has been replaced by KVL for multiple-loop circuits. The more complete versions of KCL and KVL are applied to both DC and AC circuits.

The majority of books on circuits focus on circuits that are "possible, meaning they don't violate KVL and KCL. So as to better acquaint students with circuits, it is proposed here that they be shown "impossible" circuits also.

Impossible circuits were shown in a book [8], and in the class, so as to inform students about short-circuits, and to improve their insight and conception.

These practices suggested in this paper were implemented in textbooks by the author, and in the classroom.

The concepts of this paper were practiced in the classroom by the author for many years, for three courses: one for DC circuits, one for AC circuits, and one for Systems and control. According to all indications, these methods led to better student performance and insight than from the same courses taken elsewhere. According to other instructors and students, these practices were found to better prepare students for subsequent courses.

References

[1] P. Feldmann, R. A. Rohrer, "Proof of the number of independent Kirchhoff equations in an electrical circuit," Circuits and Systems, IEEE Transactions on , vol.38, no.7, pp.681-684, Jul 1991.

[2] A. Premoli, "Revisited topology of Kirchoff's circuits," Education, IEEE Transactions on , vol.32, no.3, pp.298-304, Aug 1989

[3] Yong Wang, D. Gope, V. Jandhyala, C. –J, R. Shi, "Generalized Kirchoff's current and Voltage law formulation for coupled circuit-electromagnetic Simulation with surface Integral equations," Microwave Theory and Techniques, IEEE Transactions on , vol.52, no.7, pp. 1673- 1682, July 2004.

[4] J. Gabelli, G. Fève, J.-M. Berroir, B. Plaçais, A. Cavanna, B. Etienne, et. al., "Violation of Kirchhoff's Laws for a Coherent RC Circuit." Science, Vol. 313 no. 5786, pp. 499-502, 28 July 2006:

[5] F. R. Quintela, R. C. Redondo, N. R. Melchor, M. Redondo, "A General Approach to Kirchhoff's Laws," Education, IEEE Transactions on , vol.52, no.2, pp.273 - 278, May 2009. .

[6] A. H. Zemanian, "Nonstandard electrical networks and the resurrection of Kirchhoff's laws," Circuits and Systems I: Fundamental Theory and Applications, IEEE Transactions on, vol.44, no.3, pp.221,233, Mar 1997

[7] S. Toumpis, S. Gitzenis, "Load Balancing in Wireless Sensor Networks using Kirchhoff's Voltage Law," INFOCOM 2009, IEEE , vol., no., pp.1656,1664, 19-25 April 2009

[8] S. Khan, D.C. Circuits and Transients, 4th Edition, ISBN 978-984-33-3560-9, May 2012.

[9] S. Khan, A. C. Circuits, 2nd Edition, ISBN 978-984-33-5146-6, Jan 2013.

[10] S. Khan, Systems and Control, 3rd Edition, ISBN: 978-984-33-3561-6, February 2013.

Detection of Human's Focus of Attention using Head Pose

Sadia Afroze
Dept. of Computer Science and Engineering
Chittagong University of Engineering and Technology
Chittagong, Bangladesh
e-mail: sadiaafroze55@yahoo.com

Mohammed Moshiul Hoque
Dept. of Computer Science and Engineering
Chittagong University of Engineering and Technology
Chittagong, Bangladesh
e-mail: moshiulh@yahoo.com

***Abstract*- We address the problem of visual focus of attention (VFOA) based on human head pose. We observed the human head pose and the object that are in front of the human. We integrate the head pose with target object to monitor human attention where the human is actually looking at. The VFOA is a vital cue for drawing human attention and establishing human interaction. As it assists to understand what the person is doing and it notifies addressee-hood (who is seeing at whom). In Human Robot Interaction approach it is helpful as it attracts and control the attention of a target person depending on his or her current visual focus of attention. It is especially important to monitor human attention based on head pose as the person who are disabled and cannot speak but express their needs in terms of head pose. For recognizing a human VFOA, we noticed their head pose because head pose is an important hint in non-verbal communications. We evaluate our system in series of experiments and observed our system performance. We found a good result and got good exactness of the system performance.**

Keywords—human computer interaction; head pose; visual focus of attention ; object detection

I. INTRODUCTION

Human attention means the concentration of mind on a single object which is expressed by some physical expression and controlled by neural activities. Attention is also referred to as the allocation of limited processing resources. Eye movements, head movements or changes in body orientation are the observable behavioral responses for attracting a person's attention. It is necessary for a system to know a user's intention and focus of attention for building expert human interfaces. Detection of such information can be utilized to build natural and intuitive interfaces as the motion of a person's head pose and gaze direction are deeply related with his or her intention and attention.

One such characteristic of interest is the gaze, which indicates where and what a person is looking at, or, in other words, what the visual focus of attention (VFOA) of the person is. So, the first step in determining a person's focus of attention and intention is to track his/her gaze. In robotics in order to provide assistance to people with disabilities it is helpful for a robot to recognize the person's head pose. The robots have the potential to assist with a wide array of tasks and activities, assist people with diverse conditions, and assist people who are in bed, in a wheelchair, or are ambulating. If there is a camera in front of the disabled people and another camera in front of the objects of the disabled person, the robot can easily identify the person's attention by integrating the person's head with the target objects. Actually it is the common nature of human that he is look at the object that he needs. So, if we can implement the system it will be helpful for the disabled person who cannot speak but can express his needs or desire by using his head pose. Moreover, in absence of high definition image, we depend on human head pose to recognize the visual focus of attention. Thus, tracking the VFOA of people could have important applications for developing ambient intelligent systems. So, in our work, we consider head pose for recognizing visual focus of attention. Due to the physical placement of the VFOA targets, the identification of the VFOA can only be done using complete head pose representation. Here we use webcam for tracking the head of the person and from which we will estimate the head pose especially yaw pose of the person. Next, we will track the object which is in front of the person by another camera. Finally we will integrate this for getting visual focus of attention.

II. RELATED WORK

In HRI and HCI contexts, many conversational systems need VFOA information for analyzing and performing necessary interactions. Most works use either sensor-based or high definition image approaches which are not usually applicable or helpful for interaction with robots for recognizing VFOA. In the previous work [1] they proposed an intelligent robotic method of attracting a target person's attention in away congruent to satisfying social requirements. The used HOG pattern features and an SVM classifier for recognize VFOA. In the previous work [2] they used 3D geometric modeling for head pose estimation. Based on two human eye-areas, they model a pivot point using distance measure devised by anthropometric statistic and MPEG-4 coding scheme. But this 3D approach is quite complicated and complex rather than 2D geometric modeling. In the previous work [3] they employ neural networks to estimate a person's head pose from camera images, and a probabilistic model to recognize interesting targets in the scene based on the observed head pose. In the previous work [4], Hidden Markov Model (HMM) used to recognize VFOA and introduce the standard gaze model. They used both supervised and unsupervised learning approach. They mainly tried a improved VFOA recognition. But in practice, to set such a value might be difficult and complicated. Because in the robot interaction application, the same strategy does not produce good result in all condition. Asteriadis et al. [5] , information from head rotation and eye gaze are used. The authors use Bayesian modality fusion of both local and holistic information, in estimating head pose, while for eye gaze they use a methodology that calculates eye gaze directionality, removing the influence of head rotation. In R. Stiefelhagen et al. [6] use low-cost camera images to estimate visual focus of attention using head rotation, as well as fuzzy fusion of head rotation and eye gaze estimates, in a fully automatic manner. However, in their proposed approach [7], the user needs to maintain a frontal pose to the camera at start-up and they do not address finding appropriate mappings between 2-D projections and head/eye gaze analysis to certain points on a target plane. In contrast to previous works we would like to focus on recognizing VFOA in real time.

III. PROPOSED FRAMEWORK

In our work, we recognize the visual focus of attention of human in real time using head poses of human. To detect the visual focus of attention, first of all we track the human face and estimate its pose. The system detects the object from multiple objects corresponding to human head pose. That means the person where he is looking is detect by the system through integrating head pose and target object. Thus the system shows the visual focus of attention of human. This task is very important for a disabled person.

We also all detect the objects that are in front of the person. We use multiple object detection algorithms using shape of the object. We also number the object so that it can easily recognized by the user and it is helpful to count the number of object. Fig. 1 Illustrates the schematic diagram of propose system methodology.

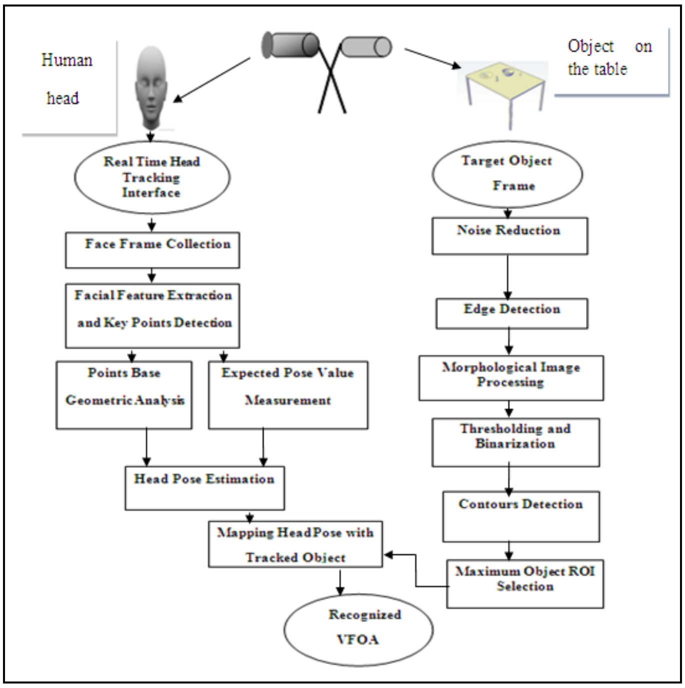


Fig. 1: Schematic diagram of the proposed system

A . Tracking the Human Head

The first step of our work is to track human face. Here we track human face using an open source named Stasm (standard active shape model). We use the Stasm 4.1.0 in our work. Stasm employ the OpenCV frontal face detector. The face should be at least a quarter of the image wide. Stasm does not use color information, i.e., it internally converts the image to monochrome before searching for landmarks. An unused or uninitialized landmark has a position in a shape matrix with both x and y equal to 0. The x position of a valid landmark that happens to be at [0,0] is thus jittered to 0.1 (a one tenth of a pixel offset). Unused landmarks are for more esoteric applications of Tasm where landmarks are synthesized or models are built with different sized shapes. It is unlikely that you will need unused landmarks when creating your own shape files. Fig. 2 represents the overview of the face tracking algorithm.

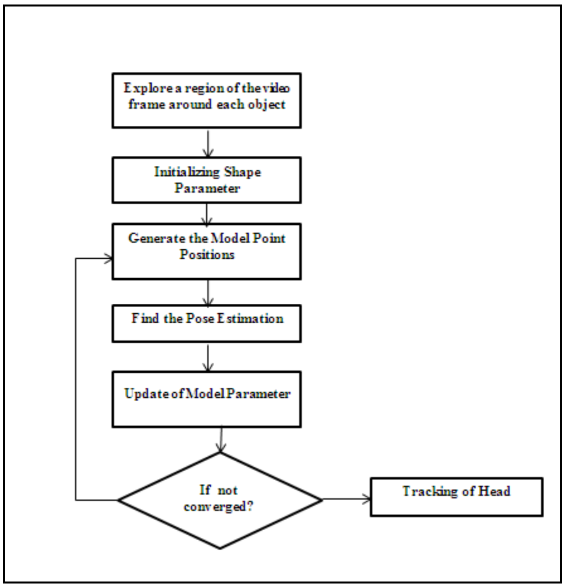


Fig. 2: Overview of face tracking process

Suppose now we have s sets of points x_i which are aligned into a common coordinate frame. These vectors form a distribution in the 2n dimensional space in which they live. If we can model this distribution, we can generate new examples, similar to those in the original training set, and we can examine new shapes to decide whether they are plausible examples. Approximate any of the original points using a model with less than 2n parameters. If we apply a PCA to the data, we can then approximate any of the training set, x using

$$x = \bar{x} + b \qquad \textbf{(1)}$$

Where $\bar{x}$ is average of all training sets, b is n dimensional vector. The vector b defines a set of parameters of a deformable model. By varying the elements of b we can vary the shape, x.

B. Head Pose Estimation of Human

After tracking the human face, we estimate human head pose. We have performed point base geometrical analysis and expected pose value measurement to estimate the head pose of the human. After performing these two measurements, we estimate the head pose of the human head to know where the visual focus of attention of him/her.

- **Expected Pose Value Measurement**: The expected value of a random variable is intuitively the long-run average value of repetitions of the experiment it represents. Suppose random variable X can take value x_1 with probability p_1, value x_2 with probability p_2, and so on, up to value xk with probability pk. Then the expectation of this random variable X is defined as:

$$E[X] = x_1p_1 + x_2p_2 + \ldots + x_kp_k \qquad (2)$$

To find the expected value of the head pose, we train our system by establishing three models. The first model is for left head pose, the second model is for front head pose and the third model is for right head pose. We observe the three models and find the threshold value for each model which was a floating point value. We observed the facial landmarks that are obtained from the face tracking module. We separate out the left side points from the right side points. We found 13 significant points on left side of the head and 13 significant points on right side of the head. Then we calculate the expected value by the equation 2. The value of current head pose value is compared with each head pose model and if it is nearest to the any model, it is said to the corresponding head pose. Thus we calculate the expected head pose of the head.

- **Point Based Geometric Analysis**: After calculating the expected value of the head pose we calculate the point bases geometric analysis of the head pose. In the learning part we first build up the point relationship model and we got some significant change in the points of the head poses. We observed that the distance between the point of eye center and nose tip is always change in the different head pose. We found out these points and calculate the Euclidian distance between them. The formula of Euclidian distance is:

$$D1\ (p, q) = \sum_{i=1}^{n} |pi - qi| \qquad (3)$$

After combining these two approaches we measure the head pose of the human.

C. Detecting the Objects in Front of the Human

In this work we detect the objects based on shape and edge. After detecting the edge we need to focus the detected object .We represent the detected objects by a rectangle. After detecting all the objects we only detect the target object. But then the bounding rectangle is on the only target objects and it represents the target object where the visual focus of attention of the person was. Fig. 3 shows the object detection algorithm.

Then we detect the edge of the object from image. Here we use Sobel edge detector for detecting edge. After detecting edges we perform morphological image processing. We use this as it process images based on shapes and apply a structuring element to an input image and generate an output image. We use erosion and dilation as structuring element. After performing morphological operation we apply thresholding and binarization of input frame.

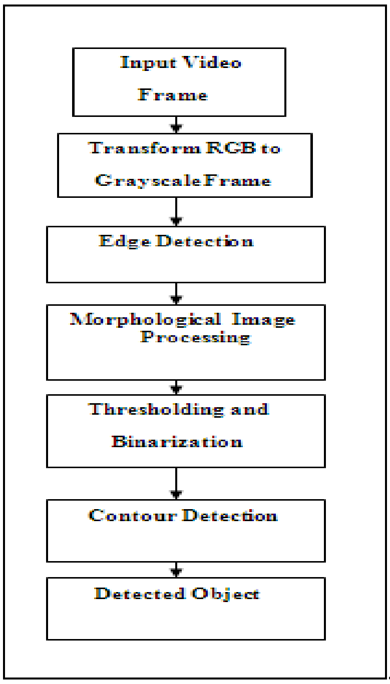


Fig. 3: Flow diagram of object detection algorithm

After this we find out the contour (boundary of shape) of the object it is like finding out the white object from black background. Each contour is stored as a vector of points. After finding the contour we estimate the bounding box of the target objects. We estimate it by analyzing the contour area of the contour. We estimate a fixed value from various shapes of the objects and compare the contour area with this value. The contours that are satisfy this constraint is considered as target objects. We represent the object by a bounding box. We use rectangles to focus the target object. The bounding box only contains the topmost coordinate, width and height of the rectangle. Thus we detect multiple objects in front of the human.

D. Integrating Head Pose with Detected Object to Recognize the VFOA

After tracking object and face, the system integrate the tracked object on the corresponding head pose. For mapping the head pose with the corresponding object the system first tracked all the objects in front of the human. Then on the basis of the head pose it recognizes the object that the person is looking at. Such as if a human moves head on the left side; the system tracked only the object left side of the person and shows that the person’s attention is on that object. On the other hand if a human moves head on the right side; the system tracked only the object right side of the person and shows that the person’s attention is on that object. Besides this, the system also shows that if the human face is frontal face, the system integrates frontal pose with middle object and indicate that his attention is in front of the middle object. Actually the integration is done on the basis of head pose and its corresponding tracked object. So we can see that we have two parameters for mapping the head pose with target object and we can represent it as follows:

VFOA= f (H, O)

Where, VFOA= Visual focus of attention, H= Head pose of the person and, O= Object where the person is looking at.

The head pose of a person is intersects with the corresponding object. We represent the target object in a rectangle. But when we integrate face tracking module with the object detection module, only the rectangle is placed only in the object with corresponding head pose. We draw a straight line from the nose tip of the person. We consider is an initial head rotation which we represent by another straight line. When the angle between initial head rotation and original head rotation zero we map this head pose with the middle object. Again there is a horizontal line from the boundary box of the object. If the two lines intersect perpendicularly then we can say that the attention of the person is on that object. If we consider the line from nose tip is y_1 and the line on the boundary box of the object is y_2, then we find out perpendicular condition using the following formula:

$$\tan(\theta)=\mathrm{mod}((m_1-m_2)/1+m_1*m_2) \quad (4)$$

Where m_1 is the slope of line that is drawn from the nose tip and m_2 is the slope of the line that is drawn from bounding box. Fig. 4 shows the mapping of head pose with target object.

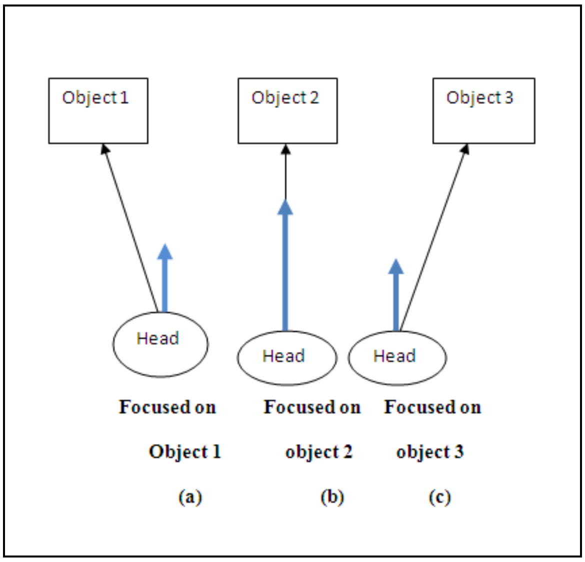


Fig. 4: Mapping of head pose with three static objects

The above shows the blue arrow represents the person's upper body resting orientation thus his or her initial head rotation. In Fig. 4 the observed head orientation (black arrow) shows the head angle towards target object. In Fig. 4(a) indicates the attention is toward the left object as it is less than the initial head pose or less than 90^0 in the x-axis. In Fig. 4(b) the observed head orientation (black arrow) is equal to the initial head pose or equal to the 90^0 in x axis and indicates the attention is toward the middle object. In Fig. 4(c) the observed head orientation (black arrow) is equal to the initial head pose or greater than to the 90^0 in the x-axis and indicates the attention is toward the right side object.

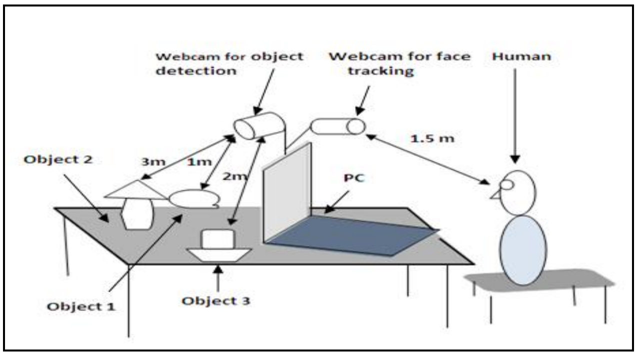
IV. EXPERIMENTAL RESULTS

To analyze the face tracking module accuracy we emphasize on quantitative approach. We worked with 10 people in the real environment to run this experiment. The average age of participants are 23 years (SD= 4.50). To evaluate the system, we told the participants about our experiment and also described them what they should do. Participants were interacted with the system one by one. Each participant attended in a two steps experiment.

A. Experimental Setup

The first phase of the implementation for this system is to establish a setup environment with two webcam. One camera can detect the human face and another can track the objects in front of the human. Fig. 5 shows the experimental set of our system.

(a)



(b)

Fig. 5. (a) Participant interacting with the proposed system (b) Experimental setup of the sysytem.

We test the system performance using different types of the object and also detect the visual focus of attention by different person. We asked the person to take sit before the webcam and set all the objects in front of the human. Then we evaluate the system performance. We can also set this webcams into a tripod to test our system performance.

B. Evaluation Method for Face Tracking Module

We positioned the participant in front of the camera within the viable range and more than 10 trials have been given by each of the participants. We build 640×480 size image of the different person of three head poses which is used as training set of the face tracking module performance. To evaluate the system, we told the participants about our experiment and also described them what they should do. The qualitative evaluation of the module, which is performed with the experimental data, is given below.

TABLE I: PERFORMANCE EVALUATION OF FACE TRACKING MODULE

	No of Trials	Tacked Face	Deviation	Accuracy (%)	Error Rate (%)
P1	15	14	1	93.33	6.66
P2	13	11	2	84.61	15.38
P3	17	15	2	88.23	11.76
P4	12	11	1	91.67	8.33
P5	19	18	1	94.73	5.26
P6	10	7	3	70	30
P7	18	15	3	83.33	16.67
P8	11	11	0	100	0
P9	17	16	1	94.11	5.88
P10	14	13	1	93	7

Graphical representation of the experiments is given below in Fig. 6.

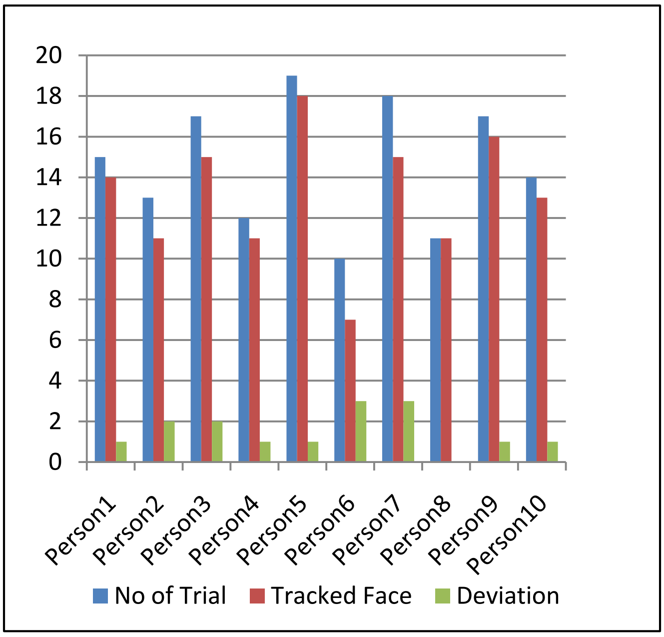


Fig. 6: Performance analysis of face angle estimation module

C. Evaluation Method of Object Tracking Module

In Fig. 7, we evaluate the object tracking module based on ROC (Receiver Operating Characteristic curve) analysis shown in.

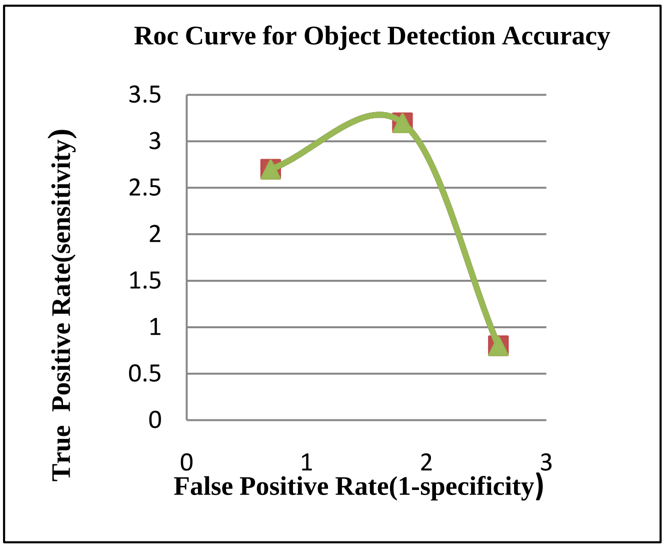


Fig. 7: ROC curve of object detection module

Fig. 7 represents a graphical plot that illustrates the performance of object tracking module. The curve is created by plotting true positive rate and false positive rate of the detected object and ground truth object.

D. Overall System Performance Analysis

We observed the performance of face tracking module and object tracking module. Then each of modules is combined to detect the visual focus of attention detection. Prior to that some important factors (i.e. face tracking of human, head pose estimation, object detection) are checked to evaluate the overall system performance. Also these values are compared to the physical values to find the standard deviation. Table II shows the data table to determine system accuracy.

TABLE II: DATA TABLE TO DETERMINE SYSTEM ACCURACY

Head Pose	No. of Trials	No. of Head Pose	Detected object with corresponding Head Pose	Error
Left	10	20	15	5
Front	16	30	20	10
Right	8	10	10	0

The graphical representation of the system accuracy is shown in Fig. 8. We observe the correct detected object of corresponding head pose and represent the accuracy by a graphical representation where we plot the number of head pose and tracked object in front of the human. We tested this system by several persons and the quantitative evaluation proved that the system working quite satisfactory.

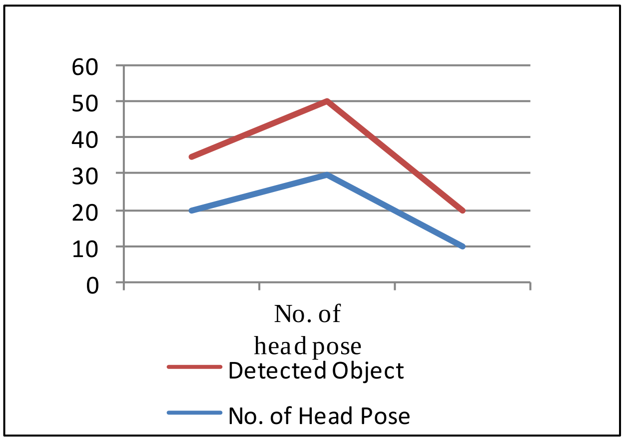


Fig.8: Graphical representation of overall system

V. CONCLUTION

Our main concern was to develop a system that recognizes the visual focus of attention of a person. We consider head pose to recognize a person attention as head pose plays an important cue for developing communication channel. In our work we addressed the visual focus of attention based on their head pose. First of all we track the human face and estimate his/her pose and finally mapped pose with corresponding VFOA targets. We use a setup of multiple camera views in order to achieve unobtrusive captures of participant's visual attention. From the estimated field of view, we deduce the most likely focus target by using an adaptive scheme of mapping head orientation to its most likely gaze angle counterpart and hence the target this person is looking at. Integration of multiple cameras will increase the field of view of the camera and can tracks the multiple human's focus of attention. This issues are left for future research.

REFERENCES

[1] R. Vertegaal, R. Slagter, G. Vander, and A. Nijholt, "Eye gaze patterns in conversations: there is more the conversational agents than meets the eyes," ACM, pp. 301–308, 2001.

[2] K. Smith, S. Ba, D. Perez, and J. M. Odobez, "Multi-person wandering focus of attention tracking," in Proc. International Conference on Multimodal Interfaces, Canada, pp. 230-240, November , 2006.

[3] J. S. Babcock and J. B. Pelz, "Building a light weight eye tracking head gear," in Proc. Eye Tracking Research and Applications, New York, NY, USA: ACM, 2004, pp. 109–114, 2004.

[4] S. Sheikhi, V. Khalidov, and J. Odobez, "Recognizing the visual focus of attention for human robot interaction," in Proc. Internal Workshop on Human Behavior Understanding, IROS, pp.120-130, October 2012.

[5] S. Asteriadis, K. Karpouzis, S. Kollias, "Visual focus of attention in non-calibrated environments using gaze estimation," Int. Journal of Computer Vision, vol. 107, No. 3, pp. 293-316, 2014.

[6] R. Stiefelhagen, J. Yang, A. Waibel, "Tracking focus of attention for human-robot communication," in Proc. IEEE-RAS International Conference on Humanoid Robots, pp. 120-130, 2001.

[7] R. J. Anderson, "Cognitive psychology and its implementation (6th ed.)", Worth Publishers, pp. 519, 2004.

[8] P. W. Rander, "Real time image based face tracking", Department of Electrical and Computer Engineering, Ms. C. Thesis, Carnegie Mellon University, pp. 130-200, 1993.

[9] F. Bashir, F. Porikli "Performance evaluation of object detection and tracking systems," in Proc. Int. conf. on CVPR, 2006.

[10] T. Cootes, "An introduction to active shape models," Model-Based Methods in Analysis of Biomedical Images" in "Image Processing and Analysis", Ed. R.Baldock and J.Graham, ,Oxford University Press, 2000, pp. 223-248.

[11] S. Aman and S. Szpakowicz, "Identifying expressions of emotion in text," Phd Thesis, School of Information Technology and Engineering University of Ottawa, Ottawa, Canada, 2007.

[12] D. D. Le, X. Wu, and S. Satoh, "Face detection, tracking and recognition for broadcast video", National Institute of Informatics, Tokyo, pp. 100-120, Japan.

[13] D. Das, Y. Kobayashi and Y. Kuno1, "Attracting attention and establishing a communication channel based on the level of visual focus of attention", vol. 6, no.2, pp. 250-280, 2013.

[14] M. S. L. khan, S. U. Rehman, Z. Lvchi, H. Li, "Geometric head pose estimation using monocular camera," in Proc. International Conference on Intelligent System and Image Processing, pp. 149-153, 2013.

[15] J. Yang, A. Waibel, "Tracking focus of attention for human-robot communication," in Proc. IEEE-RAS International Conference on Humanoid Robots, 2001.

[16] Y .S. Waibel and R. F. Karpouzis, "Object-based visual attention for computer vision", Artificial Intelligence, vol. 146, no. 1, pp. 77–123, 2003.

ID	Name
3	Abidur Rahaman, Md. Shariful Alam and ZH Mozumder
4	Tasfia Tasbin, Sabrina Abedin, Piyas Chowdhury, Md. Azad and Quazi Delwar Hossain
5	M. Bhuiyan, D. A. Firdhaus and Kh. Arif Shahriar
7	Rebeka Sultana, Samira Muntaha, Farhana Sarker, D. M. Anisuzzaman and Khondaker A. Mamun
9	Muhammad Mamunur Rashid, Rashed Mustafa and Mohammad Sanaullah
12	Mehedi Mahmud and Muhammad F. Mridha
13	Nahid Akhter and Amitabha Chakrabarty
14	Sathi Rani Mitra and Md. Mahbub Hasan
15	Mohammad Reza Rostami, Norbik Bashah Idris and Zuraini Ismail
19	Dhawa Sang Dong and Anand Gachhadar
21	Md. Nasir Uddin, M M Rashid, SZ Ahmed, Sultan Mahmud, NA Nithe and JI Rony
24	Md. Nasir Uddin, M M Rashid, SZ Ahmed, Sultan Mahmud, NA Nithe and JI Rony
29	Rupa Barua, Shauvik Shadman and Amitabha Chakrabarty
32	Md. Zulfikar Alom, Tapan Kumar Godder and Mohammad Nayeem Morshed
35	Mohammad Shoaib Shahriar, Md Shafiullah, Mohammed Afzal Asif, Md Mahmudul Hasan and Ahme
36	Imam Ush Shaheed, Sadia Sabrin Nodi and Muhidul Islam Khan
38	SM Tasdid Swad, Md. Rokebul Islam, Sadia Intesar and Jia Uddin
40	Rakibul Hasan, Jahedul Islam, Rashed Md. Murad Hasan and Adhri Nandini Paul
41	Md. Khaliluzzaman, Tanjim Mahmud, Rezaul Karim and Deepak Kumar Chy
42	Md. Kamrul Hasan, Shantanu Sen Gupta, S. M. Hasnat Ullah and Mohiuddin Ahmad
43	Mohd. Ghazali Hamza, Liza.A.Latiff, Yusaini H.Mohamed, Yusnaidi Md Yusof and M.F. Camara
44	Md. Selim Morshed, Monir Hossen, Mohammad Mahbubur Rahman and Masanori Hanawa
46	Uzzal Kumar Prodhan, Muhammad Zahidur Rahman and Israt Jahan
47	Abhijit Chowdhury and Asaduzzaman
48	Anamika Dey, Tamal Sarkar, Md. Arif Ullah and Nasrin Nahar Kamarul Ariffin Abdul Basit
49	Mohammad Obaidur Rahman and Asaduzzaman
50	Shadman Fahim, Shehabul Hossain, Gulshan Jubaed Prince and Jia Uddin
51	S.M. Mohi-Us Sunnat and Amitabha Chakrabarty
52	Golam Moktader Daiyan and Fahmida Akter
53	Azuan Ahmad, Norbik Bashah Idris and Mohd Nazri Kama
55	Mahdi Zareei, A.K.M. Muzahidul Islam, Nafees Mansoor and Sabariah Baharun
56	Md. Nazmul Hossain, Kazi Mozaher Hossein, Rajib Chakrabarty, Hiroshi Okajima, Hironobu Kitaoka ar
57	Md. Tahajul Islam, Mehedi Mahmud and Muhammad F. Mridha
58	Mohammad Shoaib Shahriar, Farhan Ammar Ahmad, Ibrahim Omar Habiballah, Mohammed Afzal As
59	Md. Habibur Rahman and Md. Mamunoor Islam
60	Md. Moniruzzaman Monir, Nahyan Ebn Hashem, Md. Nafis Hasan Siddique, Afsana Pervin Tanni and
61	Brian Sumali, Haslina Sarkan, Nozomu Hamada and Yasue Mitsukura
63	Md. Taufeeq Uddin, Mohammad Shahadat Hossain, Muhammed Jamshed Alam Patwary and Tanvee
64	A.Q.M Shaiq, Safwan Bin Alam and Md. Samiul Alam
65	Yusnaidi Md Yusof, A.K.M. Muzahidul Islam, Sabariah Baharun, Mohd Ghazali Hamza, Noor Azurati
66	Md. Khaliluzzaman and Deepak Kumar Chy
67	Abu Sayed Chowdhury, Sadia Kamal and Nusera Tasrin
70	Rabiul Islam Jony, Abiduzzaman Rahat and Musfiqur Rahman
71	Supratip Ghose
72	S. Paul, T. Sultana and M. Tahmid
73	Md. Rakin Sarder, Feroz Ahmed and Tahsin Ferdous Ara Nayna

74	Sayeda Fauzia, Shajeda Khanam, Sanjida Sharmin and Md. Khaliluzzaman
76	Muhammad Kamal Hossen and Afiya Ayman
79	Pranab Kumar Dhar, Sulogna Chowdhury and Lamia Alam
80	Faisal Ahmed, Fatema-Tuj-Johora and Mohammad Sanaullah Chowdhury
82	Mohammod Abul Kashem, Md. Hanif Seddiqui, Nejib Moalla, Aicha Sekhari and Yacine Ouzrout
83	Md. Hanif Seddiqui, Abdullah Al Mohammad and Nowshed Chy
84	Kazi Mozaher Hossein, Md. Nazmul Hossain, Ashir Ahmed, Fumihiko Yokota, Hironobu Kitaoka, Hiros
86	Emranul Haque and Feroz Ahmed
89	Md. Hayder Ali and Md. Saiful Islam
90	Md. Alamgir Hossain
91	Mohammed Saifuddin Munna, Bijoy Kumar Tarafder, Md. Golam Robbani and Tuton Chandra Mallicl
92	Sabrina Tarannum, Mohammed Saifuddin Munna and Mohammad Ariful Islam Bhuyan
93	Kanis Fatima and Md. Saiful Islam
97	Saniat Ahmed Choudhury and Mustafa Habib Chowdhury
98	Nusrat Jahan Farin, Atiqur Rahman, Nafees Mansoor and Sazzad Hossain
100	Shad Muhammad, Ghani ur Rehman, Tariq Usman, Fawad Ta., Atiqur Rahman and Asim Zeb
102	Nasimul Hasan and Risul Islam Rasel
103	Md. Mahamudul Hasan, Md. Yeasin Arafat, Mahtab Murshed and M. Abdur Razzak
104	Mazharul Islam and Nihad Karim Chowdhury
105	Md. Faisal Alam, Rashik Hafiz, Mohammad Sanaullah Chowdhury and Nihad Karim Chowdhury
108	Golap Kanti Dey, Kazi Tanvir Ahmmed and Rubell Sen Goopta
109	S. M. Nadim Uddin, Nafees Mansoor and Sazzad Hossain
111	Asim Zeb, Javed Bangash, A.K.M Muzahidul Islam, Sabariah Baharun, Atiqur Rahman and Yoshiaki Ka
112	Afsana Haque and Mohammed Moshiul Hoque
113	Md Azher Uddin, Joolekha Bibi Joolee and Shayhan Ameen Chowdhury
114	Shahriar Khan, Ziauddin Alamgir and Samina Alam
116	Mohammed Moshiul Hoque and Sadia Afroze

Title
Development of Software for the Basic Line-of-Sight Parameters Calculation
Design of a Circular Polarization Array Antenna Using Both Sided MIC Technology
Smart Window Blind Control System
Automated Credit Scoring System for Financial Services in Developing Countries
An Investigation into Various Methods of Lip Reading Systems
Augmenting ATM Security Analyzing Thermal Imaging and Voice Biometric Recognition Telecommunications
A Survey-based Study on Lip Segmentation Techniques for Lip Reading Applications
Comparison of Vocal-Tract Dynamics for Bangla Vowel and Vowel-Consonant-Vowel Sequence
A Holistic Botnet Detection Framework Independent of Botnet Protocols and Architecture
Performance Evaluation of VoIP Services over UMTS-Network with Differentiated UMTS Bearer Services
Comparative Study of Integrated Transceiver for Real Time Monitoring in Rescue Operation
Development of an Absorption Silencer for Generator's Noise Reducing
Pocket Switched Networks Routing: A Survey
Performance Analysis of Routing Protocols in Mobile Ad-hoc Network (MANET)
Comparison of Invasive Weed Optimization (IWO) and Particle Swarm Optimization (PSO) in Improving Power Systen
Context Aware Energy Allocation by Auction Based Method in Wireless Sensor Networks
A Smart LAN Infrastructure for VoIP Based Wireless Communication
Design of a Vision Based Person Following Robot
An Expert System for Clinical Risk Assessment of Polycystic Ovary Syndrome Under Uncertainty
Drowsiness Level Detection for the Protection from Accident of Intelligent Transportation System (ITS)
Statistical Prediction Model of Rain and Dust Storm Worst Month in Microwave - Millimeter Wave Band
Dynamic Hybrid Slot Size Bandwidth Allocation Algorithm for Reducing Packet Delay of Real Time Traffic in EPON
A Systematic analysis on the Telemedicine Services in Bangladesh
Selection of Optimal Number of Relays for Distributed Wireless Networks Based on Game Theory
Implementation of Improved Harmony Search Based Clustering Algorithm in Wireless Sensor Networks
Optimal Power Allocation for Multichannel Cognitive Radio Systems Using Stackelberg Game
Comparative Analysis of Protien Alignment Algorithms in Parallel Environment using CUDA
CHAAR: A Location Based Product Offer Advertisement App
An Improved Decision based Noise Reduction filter for Salt and Pepper Noise
Cloud Intrusion Detection Model Inspired by Dendritic Cell Mechanism Perspective
A Routing Protocol for Cognitive Radio Ad hoc Network
Social Adoption of ICT Based Healthcare Delivery Systems in Rural Bangladesh
Improve Payment Card Security by Adding Voice and Fingerprint Biometric Solution
Artificial Bee Colony based Optimal PMU Placement in Power System State Estimation
An Empirical Study on GSM Spectrums in Bangladesh using SDR Technology
An Efficient Magic Mirror Using Kinect
Blind Image Restoration by using PCA-Subspace Generation and Image Quality Optimization
Predicting the Popularity of Online News using Gradient Boosting Machine
Design and Performance Analysis of CNFET Oscillator and Comparison with its CMOS Implementation
An Environmental Sensing Experiment Using IEEE 802.15.4 Radio Hop of the WSN TelosB Motes
Compressed Image Transmission over AWGN Channel using DCT and Rai Sed Cosine Filter
Performance Survey of HSPA Network in Chittagong City
Big Data Characteristics, Value Chain and Challenges
Automatic Face Annotation with Face Prototype Map in Personal Photo Frame Application
Automatic Electrical Home Appliance Control and Based Brain-Computer Interfacing
Modeling a Zigbee and PLC Based Smart Energy Monitoring and Management System to Reduce Rolling Blackout in E

An Android Based Femina Security Alert System
JRanker: An Approach to Evaluate the Prestige of a Journal Using PageRank and Alexa Rank along with Impact Factor
A System to Ensure Privacy for Android Users
A Location Based Smartphone Application to Rent Private Vehicles at Real Time
Review on Telemonitoring of Maternal Health care Targeting Medical Cyber-Physical Systems
Recursive Suffix Stripping to Augment Bangla Stemmer
Towards Reducing BoP Penalty through Rural E-Commerce: Optimization of Product Delivery Mechanism
ECG Signal Based Heart Disease Detection System for Telemedicine Application
GPON Triple Play and SDH Connectivity Structure with Cost Analysis
Basic Sequential Algorithmic Scheme Based Blind Common Phase Error Compensation in OFDM Systems
Design and Implementation of a Painter Robotic Arm with Graphical User Interface
Gesture Based Wireless PC Control with Gyroscope and Accelerometer
Analysis of Chirp Induced Impairments in Fiber Optic Transmission Systems Using Various Types of Fiber
The Promise and Challenges of Enhancing Solar Cell Efficiency Using Patterned Nanostructures
WoTCoMS: A Novel Cross-Layered Web-of-Things Based Framework for Course Management System
Simulation Based Study to Present the Performance of Ad-hoc Routing Protocol
Artificial Neural Network Approach for Stock Price and Trend Prediction
Sensorless Temperature Monitoring System using GSM Module for Smart Home Applications
Phishing Websites Detection Using Machine Learning Based Classification Techniques
Performance Analysis between Probabilistic and Decision Tree based Classification on User Knowledge Model Datase
Novel Approach of Phased Array Antenna with Beam Steering Technology for Microwave Power Transmission from S
Cognitive Radio Enabled VANET for Multi-agent Based Intelligent Traffic Management System
Network Formation and Data Centric Routing in Wireless Sensor Networks
Bangla Word Sense Disambiguation System using Dictionary Based Approach
Bangladeshi Vehicle Digital License Plate Recognition for Metropolitan Cities Using Support Vector Machine
Generalized Versions of Kirchhoff’s Laws for Students
Detection of Human’s Focus of Attention using Head Pose

3
4
5
7
9
12
13
14
15
19
21
24
29
32
35
36
38
40
41
42
43
44
46
47
48
49
50
51
52
53
55
56
57
58
59
60
61
63
64
65
66
67
70
71
72
73
74
76
79
80

82
83
84
86
89
90
91
92
93
97
98
100
102
103
104
105
108
109
111
112
113
114
116