International Journal of Computer Applications Technology and Research
Volume 5– Issue 6, 320 - 329, 2016, ISSN:- 2319–8656

Simulative Performance Evaluation of a Free Space Optical Communication Link Operating at 1550 nm using Different Modulation Formats

Salwinder Singh
Department of Electronics Technology
Guru Nanak Dev University, Amritsar
Punjab, India

Karamdeep Singh
Department of Electronics Technology
Guru Nanak Dev University, Amritsar
Punjab, India

Shivinder Devra
Department of Electronics Technology
Guru Nanak Dev University, Amritsar
Punjab, India

ABSTRACT

FSO enables to provide last mile transmission reach with some advantages such as inherent security and no electrical hazards without laying any fiber or cable. In this paper we have analyzed the performance of free space optics communication system by employing five transmitters (CSNRZ CSRZ DBNRZ MDBNRZ and MDBRZ) for different attenuation values at 10Gbps up to the transmission distance of 1100m. The effect of beam divergence, transmitter losses, receiver losses and receiver aperture diameter is also calculated on the performance of proposed hybrid modulation format based free space optical communication system.

Key words: FSO; CSNRZ; CSRZ; DBNRZ; MDBNRZ; MDBRZ;

I. INTRODUCTION

Free space optics is a wireless communication technology which utilizes light for transmission of data through the air in the similar manner as the fiber optics uses a fiber cable. Free space optics is having the same capabilities as that of fiber optics, but at a very lower cost and very fast deployment speed [1]. It has advantages high speed, low cost, high bandwidth, quick installation high security and also license-free longer range spectrum [2]. But in fiber case have problem that is dispersion, nonlinear impacts in a transmission line but not in the FSO channel [3-5-6] and FSO system is also severely limited by the four wav mixing effect [4]. Free space optics work on the principal of laser driven technology which use light source and detectors to transmit and receive information, through the atmosphere same as the Fiber Optic communication cable [2]. The motivation for FSO is to eliminate the cost, time, and effort of installing fiber optic cable, yet to retain the benefit of high data rates for transmission of voice, data, images and video. FSO communication can be used in building to building, ship to ship, aircraft to ground and satellite to ground. The stability quality of the link is highly dependent on atmospheric factors such as Fog, Rain, Haze and Heat [7]. FSO system consists of

an optical transceiver at the both ends to provide full duplex capability. FSO communication is not a new technology. FSO is a LOS (Line of sight) technology, where data, video and voice communication is achieved with maximum 10Gbps of data rate by full duplex [8]. An effective FSO system should have the following characteristics [8]:

(a) FSO system should have the ability to operate at higher power levels for longer distance.
(b) For high speed FSO systems, high speed modulation is important.
(c) An overall system design should have small footprint and low power consumption.
(d) FSO should have the ability to operate over wide temperature range and the performance degradation would be less for outdoor system.
(e) Mean time between failures of system should be more than 10 years.

An effective FSO system usually has the following Merits [9]

(a) FSO is a flexible network that delivers better speed than Broadband.
(b) Installation is very easy and it takes less than 30mintues to install normal locations.
(c) It has very low initial investment.
(d) It is a secure system because of line of sight operations and so no security system up gradations is needed.
(e) There is relatively high bandwidth.

The advantages of an effective FSO system are linked below [9]:

(a) Physical obstructions: -flying birds, tree and tall building can temporary block a single beam, when it appears in the LOS of transmission of FSO system.
(b) Geometric losses: - Geometric losses which can be called optical beam attenuation are induced due to the spreading of beam.
(c) Atmospheric turbulence: - It is caused by weather and environment structure.
(d) Atmospheric Attenuation: - It is the result of Fog and haze normally.
(e) Scattering: - Scattering phenomena happen when the optical beam and scatter collide.

In the section 2, literature survey is discussed. In the section 3, different modulation formats along with their optical spectrum are described. In section 4, explains simulation setup and parameters. In section 5, results have been reported for various formats. In section 6 conclusions are made.

II. LITERATURE SURVEY

Ajay K. Sharma et al. (2009) studied robustness of various modulation formats at 40Gbps. The performance is categorized using Q-factor. They investigated non linearity and noise show robustness up to 450 km at 40Gbos. At high rate, CRZ show better results than NRZ, RZ and CSRZ [10].

Jagjit Singh Malhotra et al., (2010) investigate the Performance analysis of NRZ, RZ, CRZ and CSRZ data formats in 10Gbps. In this paper investigate the performance of NRZ, RZ, CRZ and CSRZ data formats analyzed on the basis of bit error rate (BER), Q2 (dB), OSNR, eye opening performance metrics. The results show that CRZ and CSRZ modulation format is perform better as compared to NRZ and RZ. The CSRZ has optimal performance according to performance metrics [11].

Malti et al. (2012) studied advance modulation format at different bit rates and observed that MDBRZ show better performance as compared to DBRZ and CSRZ at high bit rates but at 2.5Gbps CSRZ is better than DBRZ and MDBRZ [12].

Jitendra Singh et al., (2013) investigate different modulation format based on the performance analysis of free space optical communication system. In the FSO network some factor play important role i.e. bit error rate (BER), Q factor, forward error correction (FEC), attenuation, absorption, Scattering and scintillation. In this paper investigate the impact of different direct and external modulation formats i.e. RZ, CRZ, CSRZ and NRZ on free space optical communication system. The external modulation has better performance as compare to direct modulation because direct NRZ spectrum has a strong carrier component compared to external modulated NRZ. The simulation results prove that RZ modulation format is best for long distance, but is complex and costly. Where NRZ is used for short distance and it is less complex, cheaper in comparison to RZ [13].

Jun He et al., (2014) discussed the survey on recent advances in optical communications. The FSO is used in various applications. In this paper investigate the overview of recent research in optical communications and focus on the topics of modulation, switching, add-drop multiplexer, coding schemes, detection schemes, orthogonal frequency-division multiplexing, system analysis, cross-layer design, control and management, free space optics, and optics in data center network. The author aim is provide the knowledge about the advances in optical communications. Hence from this survey conclude that optical communication plays important role in telecommunication and data center communications [14].

Mazin Ali et al. (2014) analyzed data rate for FSO system and showed that the data rate decreases with increasing divergence angle and link distance [15].

Ajay K. Sharma et al. (2009) studied robustness of various modulation formats at 40Gbps. At high rate, CRZ show better results than NRZ, RZ and CSRZ. Jagjit Singh Malhotra et al., (2010) investigate the Performance analysis of NRZ, RZ, CRZ and CSRZ data formats in 10Gbps. The results show that CRZ and CSRZ modulation format is perform better as compared to NRZ and RZ. The CSRZ has optimal performance according to performance metrics. Malti et al. (2012) studied advance modulation format at different bit rates and observed that MDBRZ show better performance as compared to DBRZ and CSRZ at high bit rates but at 2.5Gbps CSRZ is better than DBRZ and MDBRZ. Jitendra Singh et al., (2013) investigate different modulation format based on the performance analysis of free space optical communication system. The simulation results prove that RZ modulation format is best for long distance, but is complex and costly. Where NRZ is used for short distance and it is less complex, cheaper in comparison to RZ. Jun He et al., (2014) discussed the survey on recent advances in optical communications. The FSO is used in various applications. Hence from this survey conclude that optical communication plays important role in telecommunication and data center communications. Mazin Ali et al. (2014) analyzed data rate for FSO system and showed that the data rate decreases with increasing divergence angle and link distance. It has been observed from the study of available literature that most of researchers have contributed towards enhancement of fiber optic communication links by different modulation formats, whereas none of the precision researchers have investigated the performance of free space optical communication link by employing different modulation formats.

III. DIFFERENT MODULATION FORMATS

1) **Carrier Suppressed Non return to Zero (CSNRZ) Format: -** CSNRZ format has narrower pedestal shape of the optical spectrum than the conventional RZ format. Fig-(1a) shows the schematic diagram of CSNRZ transmitter. In it the NRZ signal after MZ modulator goes through the phase modulator driven by analog sine wave generator at the frequency equal to half the

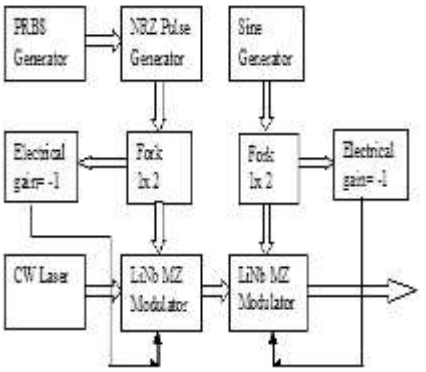
bit rate. That will introduce zero phase-shift between the two adjacent bits and the spectrum will be modified such that the central peak at the carrier frequency is suppressed as shown in the Fig-(1b).

2) **Carrier Suppressed Return to Zero (CSRZ) Format: -** Fig-(2a) shows the schematic diagram of CSRZ transmitter. In it the RZ signal after MZ modulator goes through the phase modulator driven by analog sine wave generator at the frequency equal to half the bit rate. That will introduce zero phase shift between the two adjacent bits and the spectrum will be modified such that the central peak at the carrier frequency is suppressed as shown in the Fig-(2b) [16].

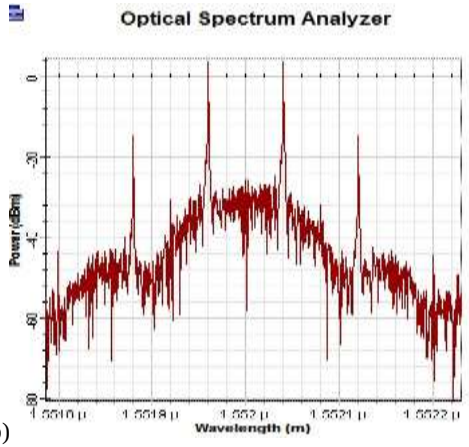
3) **Duo Binary Non-return to Zero (DBNRZ) Format: -** Fig-(3a) shows the diagram of DBNRZ transmitter. The Duo binary was generated by first creating NRZ duo binary signal using a duo binary precoder, NRZ generator and a duo binary pulse generator. The generator drives the first MZM and then the cascades this modulator with a second modulator that is driven by a sinusoidal electrical signal with the frequency on bit rate and phase= -90. The duo binary precoder used here is composed of an exclusive-or gate with a delayed feedback path and frequency spectrum shown in Fig-(3b) [15].

4) **Modified Duo-Binary Non return to Zero (MDBNRZ) Format: -** Fig-(4a) show the schematic diagram of MDBNRZ transmitter are called carrier suppressed duo binary format. MDBNRZ was generated by first creating an NRZ duo binary signal using a delay and subtract circuit that drives the first MZM and then concentrating this modulator with a second modulator that driven by a sinusoidal electrical signal with the frequency equal to the bit rate and phase equal to -90. The generation of MDBNRZ signal is same as the DBNRZ signal except the delay-and-add circuit is replaced by a delay-and-subtract circuit. In the duo binary signal used earlier where the phase of bits '1' are modified only after a bit '0' appear where as in the modified duo binary signal the phase is alternated 0 and 180 for the bits '1'. Also optical signal spectrum shown in Fig-(4b) that carrier of the duo binary signal has been-suppressed.

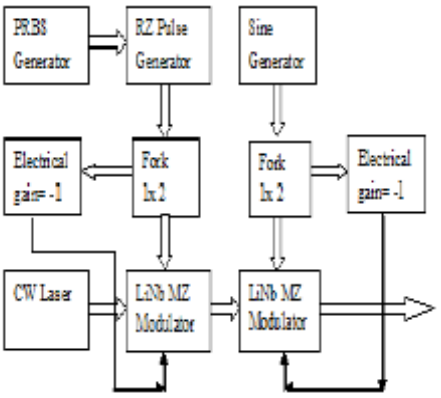
5) **Modified Duo-Binary Return-to-Zero (MDBRZ) Format:-**Fig-(5a) show the schematic diagram of MDBRZ transmitter format. MDBRZ was generated by first creating an RZ duo binary signal using a delay and subtract circuit that drives the first MZM and then concentrating this modulator with a second modulator that driven by a sinusoidal electrical signal with the frequency equal to the bit rate and phase equal to -90. The generation of MDBRZ signal is same as the DBNRZ signal except the delay-and-add circuit is replaced by a delay-and-subtract circuit. In the duo binary signal used earlier where the phase of bits '1' are modified only after a bit '0' appear where as in the modified duo binary signal the phase is alternated 0 and 180 for the bits '1' frequency spectrum shown in Fig-(5b) [16].



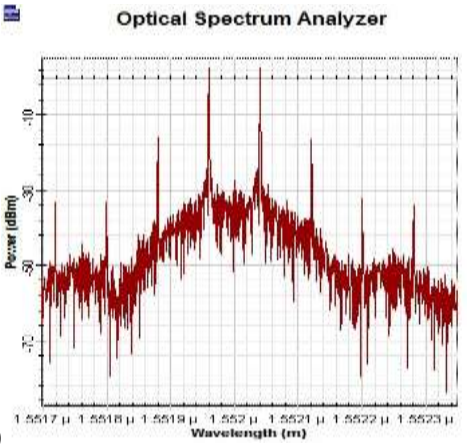
1a)



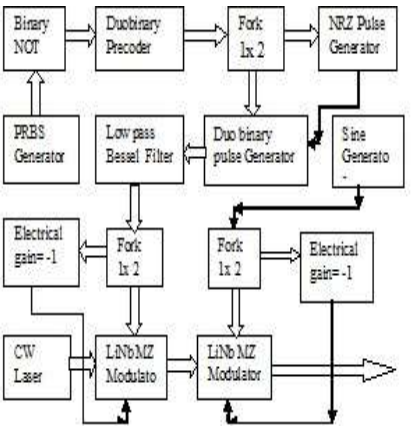
1b)



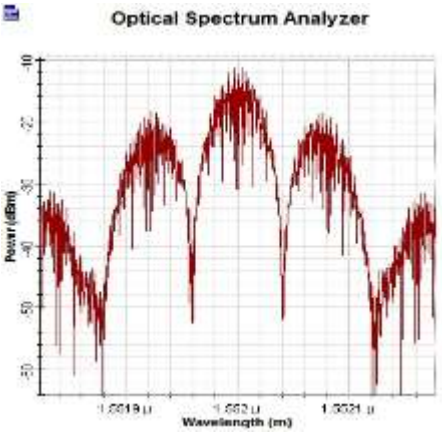
2a)



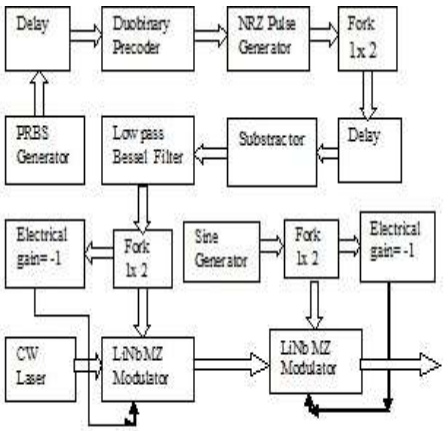
2b)



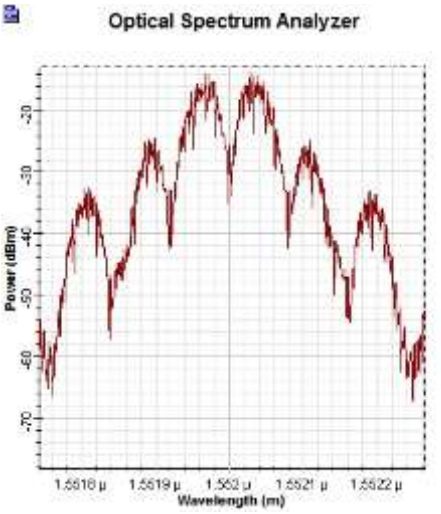
3a)



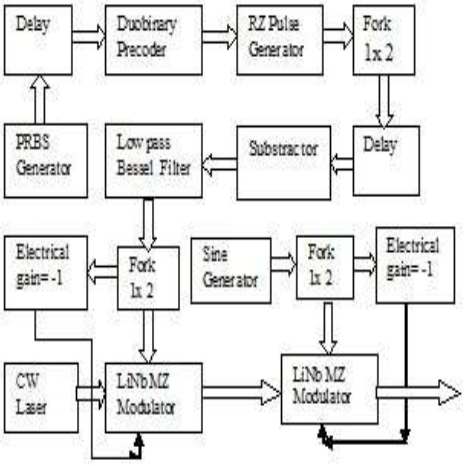
3b)



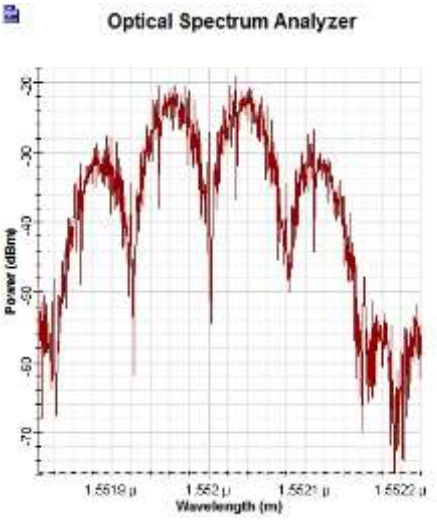
4a)



4b)



5a)



5b)

Fig.1. (1a) schematic of CSNRZ transmitter, (1b) frequency spectrum of CSNRZ transmitter, (2a) schematic of CSRZ transmitter, (2b) frequency spectrum of CSRZ transmitter, (3a) schematic of DBNRZ transmitter, (3b) frequency spectrum of DBNRZ transmitter, (4a) schematic of MDBNRZ transmitter, (4b) frequency spectrum of MDBNRZ transmitter, (5a) schematic of MDBRZ transmitter, (5b) frequency spectrum MDBRZ transmitter.

IV. SIMULATION SETUP

Fig-2 shows a schematic setup of a single channel optical free space communication system operating at 10Gbps with the central frequency 1552nm.

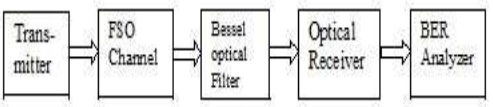


Fig-2-Block-Diagram-of-Fso

The simulation parameters used in the system model are given in Table 1.

Table-1 Simulation parameters of FSO

Parameter	**value**
Bit rate	10Gbps
Sequence length	128
Samples/bit	64
Central Frequency of channel	1552nm
Range	1100m

The simulation setup is composed of transmitter, free space optical channel and receiver. The transmitter consists of a CW laser, data modulator as shown in Fig-1 and to each output port of the CW laser a data modulator has been connected. The optical Transmitter consists of three subsystems. The first subsystem is the Pseudo-Random Binary Sequence generator. This subsystem is representing the information or data that want to be transmitted. The output from a PRBS generator is a bit stream of a binary pulse; a sequence of "1"s (ON) or "0"s (OFF), of a known and reproducible pattern. The second subsystem is very different modulation format (CSNRZ CSRZ DBNRZ MDBNRZ MDBRZ) electrical pulse generator as discussed in section3. The third subsystem in the optical transmitter is the direct modulation lasers. Laser operating wavelengths around 1550nm were developed specially for fiber optic communication because of low attenuation characteristics on this range. The free space between transmitter and receiver is considered as FSO channel which is propagation medium for transmitted light. The receiver is use to regenerate electrical signal of the original bit sequence and the modulated electrical signal as in the optical transmitter to be used for BER analysis.

V. RESULTS AND DISCUSSION

The five different modulation formats have been numerically compared for 10Gbps FSO system in terms of received maximum Q value. To analyze the system, the result of the channel has been taken. As shown in Fig-3 here in this simulation, we are varying the power from 11dbm to 20dbm at the transmitter side and Attenuation is 20dB/km and Beam divergence is 0.5mrad.With increase in the power at transmitter side, it has been observed that DBNRZ system response good as compared to other system. Also MDBNRZ have shown less-variations as we have increased the power. DBNRZ system is better response than CSNRZ, CSRZ, MDBRZ and MDBNRZ systems.

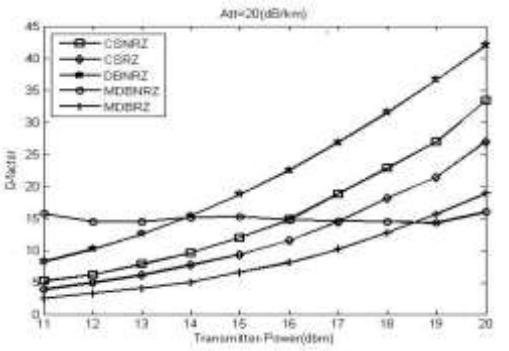


Fig-3 Plot between Q-Factor and Power

Further in order to validate the precision result that performance of DBNRZ is better, further investigation have been carried out for different value of the Beam divergence. The power is also set to 20dbm and attenuation is 20dB/km and beam-divergence varying form 0.1mrad to 1mrad; the performance of the five systems degrades as shown in the Fig-4. The performance of the DBNRZ system is better than the other systems. For beam divergence=1mrad, DBNRZ system have Q factor is up to 16.47, whereas MDBRZ system have Q factor= 5.59 and CSNRZ system have Q factor = 10.46

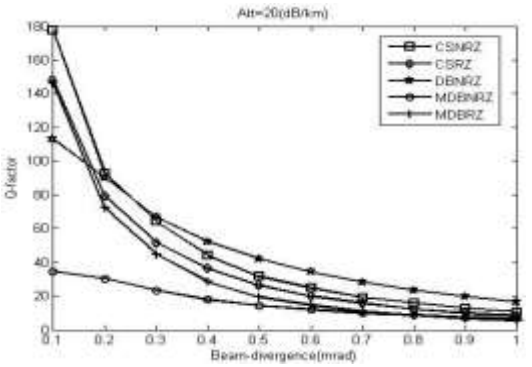


Fig-4 Plot between Q-factor and Beam divergence

For evaluating the performance, we have analyzed five transmitters for free space optics communication by using the different attenuation values.

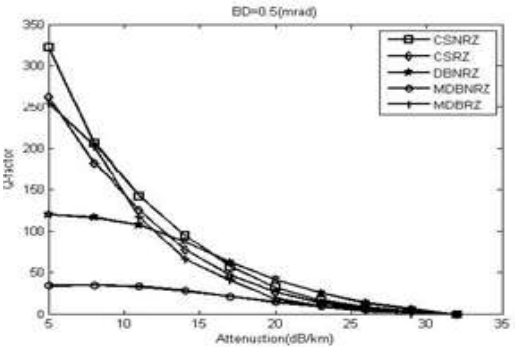


Fig-5 Plot between Q-Factor and Attenuation

Here we have taken attenuation values from 0 to 35dB/km and Beam-divergence is set to 0.5mrad. As the attenuation increases, the Q factor decreases for all system except MDBRZ system because there is very less variation. The overall performance for different values of attenuation is better in all case except MDBRZ system. In the CSNRZ system gives high Q factor value for low value of attenuation, but falls as attenuation increase as shown in the Fig-5.

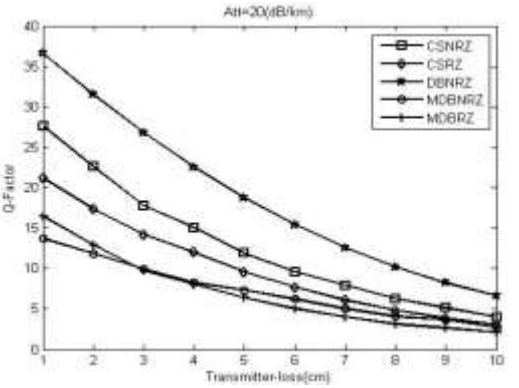


Fig-6 Plot between Q-Factor and Transmitter Loss

We have analyzed the five systems for the transmitter losses and the receiver losses as shown in Fig-6 and Fig-7.

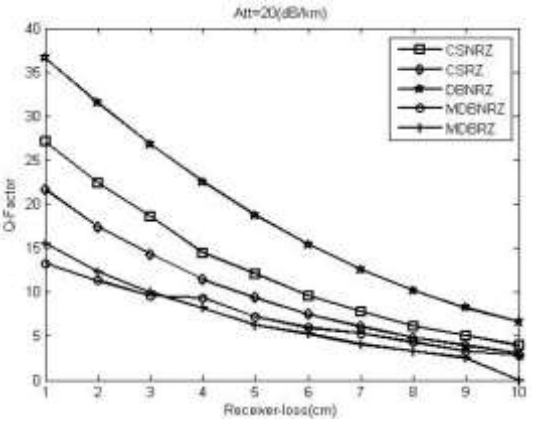


Fig-7 Plot between Q-Factor and Receiver Loss

From the results it is observed that the performance degrades for the increase in both transmitter and receiver losses. In the both case DBNRZ high Q factor, so DBNRZ system better performance as compare to the system.

We have also analyzed the system performance using receiver aperture diameter as shown in Fig-8. As we increase the receiver aperture diameter (from 25cm to 115cm and Beam-divergence is 0.5mrad) the Q factor value increases.

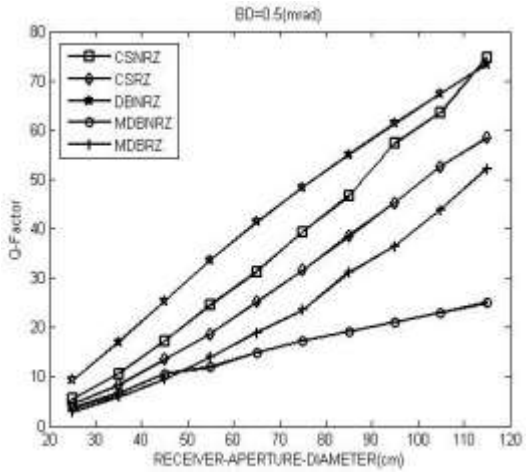


Fig-8 Plot between Q-Factor and RAD

This increased in Q factor value for DBNRZ system more than the other system. All system has almost similar response except MDBNRZ system because this system has very less variation in the Q-factor as compared to other system.

V. CONCLUSION

In this paper, we have studied the performance of five transmitters DBNRZ, CSNRZ, CSRZ, MDBNRZ and MDRZ for the free space optical communication by evaluating the performance using various parameters such as attenuation, Beam-divergence, Receiver aperture diameter, Transmitter Losses and Receiver Losses. From the simulation results, it is concluded that the DBNRZ transmitter gives the better performance for all the parameters except attenuation where CSNRZ has better results. Also it is analyzed that the performance of each transmitter degrades with the increase in Beam-divergence, Transmitter Losses and Receiver Losses as the Q factor also decreases. Finally it is concluded that the overall performance of the DBNRZ system is better than CSNRZ, CSRZ, MDBNRZ and MDBRZ transmitters.

VI. REFERENCES

[1] Kolka, Z., Wilfert, O., Kvicala, R., Fiser, O. Complex Model Of Terrestial FSO Links. Proceedings Of Spie,2007, VOL.6709,P.67091J.

[2] David, F. Scintillation Loss In Free-Space Optic Im/Dd Systems. In Lase 2004, Vol.5338. San Jose(U.sa),2004.

[3] Shivinder Devra, Gurmeet Kaur, "Different Compensation Techniques to Compensate Chromatic Dispersion in Fiber Optics", International Journal of Engineering and Information Technology, vol. 3, no. 1, pp. 1-4, 2011 (ISSN 0975-5292 (Print), ISSN 0976-0253.

[4] Karamdeep Singh, Gurmeet Kaur, Maninder Lal Singh, "A single As2Se3 chalcogenide Highly Non-Linear Fiber (HNLF) based simultaneous all-optical half-adder and half-subtracter", locate, vol. 24, pp. 56-63, April 2015.

[5] Shivinder Devra, Gurmeet Kaur, "Dispersion Compensation using Raised Cosine Filter in optical fibers", International Journal on Information and Electronics Engineering, vol. 1, no. 1, pp. 47-51, July 2011 (ISSN: 2010-3719)

[6] Shivinder Devra, Gurmeet Kaur, "Dispersion Compensation using All Pass Filters in Optical Fibers", Proc. 2011 International Conference on Information and Electronics Engineering, IACSIT press, vol. 6, no. 5, pp. 218-222, 2011.

[7] Ashish Kumar, Aakash Dhiman, Devender Kumar, Naresh Kumar, " Free Space Optical Commmunication System under Different Weather Conditions",IOSR Journal of Engineering, e-ISSN: 2250-3021, p-ISSN:2278-8719, VOL. 3, ISSUE 12, V2, PP. 52-588, 2013.

[8] http://www.laseroptronics.com/index.cfm/id/57-66.htm.

[9] Aditi Malik, Preeti Singh, "Free Space Optics: Current Applications and Future Challenges", Hindawi Publishing Corporation, Interntionl journal of Optics, vol.2015, Article ID 945483, 7 pages.

[10] Ajay K. Sharma, S.K. Wadhwa, T.S. Kamal," Robustness of NRZ , RZ, CSRZ modulation on fiber Nonlinearities and amplifier noise at 40Gbps for Long Haul Link", Optik International Journal for Light and Electron optics, vol. 120, pp. 614-618, 2009.

[11] Jagjit Singh Malhotra, Manoj Kumar, Ajay K. Sharma, Alok Kumar, "Performance evaluation of 16 channel DWDM radio-over-fiber link", International Journal for Light and Electron optics, vol. 124, pp. 4120-4122, 2013

[12] Malti, Meenakshi, Anu Sheetal," Comparison of CSRZ, DRZ and MDRZ Modulation Formats for High Bit Rate WDM-PON System Using AWG, " International journal of Emerging Technology and Advances Engineering., vol. 2, issue 6, 2012.

[13] Jitendra Singh, Naresh Kumar, "Performance analysis of different modulation format on Free Space optical communication system", International Journal for Light and Electron optics, vol. 124, pp. 4651-4654, 2013.

[14] Jun He , Robert A. Norwood , Maïté Brandt-Pearce , Ivan B. Djordjevic , Milorad Cvijetic ,Suresh Subramaniam , Roland Himmelhuber , Carolyn Reynolds , Pierre Blanche ,Brittany Lynn , Nasser Peyghambarian , "A survey on recent advances in optical communications", Computers and Electrical Engineering, vol.40, pp.216-240, 2014

[15] Mazin Ali A. Ali "Analysis of Data Rate for Free Space Optics Communication System", International Journals of Electronics & Communication Technology, vol. 5, no. 1, pp. 20-23, March 2014.

[16] Anu Sheetal, Ajay K.Sharma, R.S. Kaler, "Simulation of high capacity 40 Gb/s long haul DWDM system using different modulation formats and dispersion compensation schemes in the presence of Kerr's effect", International Journal for Light and Electron optics, vol. 121, pp.739-749, 2010.

International Journal of Computer Applications Technology and Research
Volume 5– Issue 6, 330 - 337, 2016, ISSN:- 2319–8656

Comparative Study of Lip Extraction Feature with Eye Feature Extraction Algorithm for Face Recognition

Aakanksha Agrawal
Department of CSE
RCET, Bhilai
CG, India

Abstract: In recent time, along with the advances and new inventions in science and technology, fraud people and identity thieves are also becoming smarter by finding new ways to fool the authorization and authentication process. So, there is a strong need of efficient face recognition process or computer systems capable of recognizing faces of authenticated persons. One way to make face recognition efficient is by extracting features of faces. This paper is to compare the relative efficiency of Lip Extraction and Eye extraction feature for face recognition in biometric devices. Importance of this paper is to bring to the light which Feature Extraction method provides better results under various conditions. For recognition experiments, I used face images of persons from different sets of YALE database. In my dataset, there are total 132 images consisting of 11 persons & 12 face images of each person.

Keywords: face recognition, feature extraction, lip extraction, eye extraction, GLCM, SIFT, SVM

1. INTRODUCTION

Biometrics is the study of methods for measuring unique biological and psychological characters of human being that can be used for uniquely recognizing or verifying the individual's identity.

Face Recognition is an identification procedure in which a person is verified based on human traits. Over the past years, Face Recognition gained a lot of attention in the scientific and industrial areas. A number of face recognition algorithms and their extensions, along with distinct measure of success and different accuracy level have been proposed. However, face recognition faces challenging problems in real life applications due to the variation in illumination, position, emotions and expression of the face images. Thus, still better techniques and algorithms are needed to be implemented so as to perform and create better face recognition, better face features extraction, better face detection results, etc.

Feature Extraction is a kind of process for reducing dimensionality so as to represent the interesting image parts with great efficiency as a compact feature vector [16]. This is useful in case of large images.
No one can say which algorithm is best suitable for extracting features. The algorithm selection is dependent on:

1) What exactly is the task it needs to perform?
2) Whether supervised method is needed or unsupervised?
3) Whether Inexpensive method is required or strong computational method is required?

Some numerical programming environments such as MATLAB provide techniques for feature extraction.

Lip Extraction is used as an identifier because the size of the upper and lower lips, furrows, grooves, the distances between the lines and edges vary individualistically. The uniqueness of lip has been proven by the researchers by using colour information and shape analysis. The highlighting advantage in devising the biometric system based on lip recognition is that the data acquisition and handling are simple and hence storing and processing the data becomes an unconstrained task. There are researches carried out in the direction of extracting feature from lip image by shape analysis and color analysis. However lip shape can be changed to a large extent with the change of expression and color of lip is subject to acquisition conditions. Hence, analysis and recognition of an individual by lip-shape and lip-color analysis can only work fine in a constrained scenario. However, the local feature points of a lip image can be exploited as they are invariant to illumination, rotation, scaling and other affine transforms.

2. LITERATURE REVIEW

Sambit Bakshi, Rahul Raman, Pankaj K Sa paper proposes that grayscale lip images constitute local features. The claim has been experimentally established by extracting local features applying two techniques viz. SIFT and SURF. The results obtained are enough to establish that unique local features exist in lip images through which an individual can be recognized. [India Conference (INDICON), 2011 Annual IEEE].

Ishvari S. Patel, Apurva A. Desai used Preprocessing techniques like Edge Detection by Canny Method and Height and Width comparison for Lip Contour Detection. This model works effectively and gives around 98% result for image sequences but we can still improve accuracy of result by extracting perfect lips region. [International Journal of Scientific Research (IJSR), Volume II, Issue V, May 2013].

Zeynep Orman, Abdulkadir Battal, Erdem Kemer recognised that haar-like feature approach is the most used method for face recognition projects (especially for video detection of faces and eyes) in recent years, due to be able to work in real time systems with great performance. [International Journal of Computer Science & Engineering Survey (IJCSES) Vol.2, No.3, August 2011].

Sasikumar Gurumurty, B. K. Tripathy divided methodology into: Mouth Region Localization and Key point's Extraction

and Model Fitting. In first and second steps, mouth region and key points are found by using hybrid edges, which combine color and intensity information. In third step, cubic polynomial models are fitted using key points position and hybrid edges. An automatic, robust and accurate lip segmentation method has been presented. This is considered as good result and encourage for its use combined with other biometrics systems.[I.J. Intelligent Systems and Applications (IJISA), July 2012 in MECS].

B. Sangeetha Devi, V.J.Arul Karthick used two processes for lip recognition. First, face detection by Viola and Jones algorithm. Second, lip detection by morphological operation and five various mouth corner points. Lip biometric can be used to authenticate an individual since the lip is unique. [International Journal of Advanced Research Trends in Engineering and Technology (IJARTET), Vol. II, Special Issue I, March 2015].

Tanmay Rajpathak, Ratnesh Kumar, Eric Schwartz proposed a novel technique for eye detection using color and morphological image processing. Firstly, the skin region is detected using a color based training algorithm and six-sigma technique operated on RGB, HSV and NTSC scales. Further analysis involves morphological processing using boundary region detection and detection of light source reflection by an eye, commonly known as an eye dot. This technique is found to be highly efficient and accurate for detecting eyes in frontal face images. [Florida Conference on Recent Advances in Robotics (FCRAR) 2009].

Rutuja G. Shelke, S.A.Annadate presented a novel approach for Face Recognition and Gender classification strategy using the features of lips. Here feature extraction is carried out by using Principal component analysis (PCA) and Gabor wavelet. Out of two techniques, results of Gabor filter are more accurate and fast because it is having less leakage in time frequency domain. [International Journal of Innovation and Scientific Research (IJISR), Vol 10, No.2, Oct.2014, Innovative Space of Scientific Research Journals (ISSR)].

Noopur Desai, Vidhi Dharia, Ishani Shah, Ghanshyam Prajapati attempts to detect eyes in human images and have implemented eye detection using EyeMap. The proposed approach is divided into two phases. The first phase deals with the detection of face in the image. Once the face is detected, the second phase deals with the detection of eyes in the face detected. The results of the implemented algorithm are quite accurate. [International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE), Vol. 1, Issue 3, May 2013].

Mayank Vatsa, Richa Singh, Afzel Noore proposes algorithms for iris segmentation, quality enhancement, match score fusion, and indexing to improve both the accuracy and the speed of iris recognition. A curve evolution approach is proposed to effectively segment a nonideal iris image using the modified Mumford–Shah functional. Different enhancement algorithms are concurrently applied on the segmented iris image to produce multiple enhanced versions of the iris image. [IEEE Transactions on Systems, Man and Cybernetics—Part B: Cybernetics].

5. METHODOLOGY

Himanshu Srivastava described the novel techniques developed to create an Iris Recognition System available. This paper proposed a personal identification using iris recognition system with the help of six major steps i.e. image acquisition, localization, Isolation, normalization, feature extraction and matching and also these six steps consists of minor steps to complete each step. The boundaries of the iris, as papillary and limbic boundary, are detected by using Canny Edge Detector & Circular Hough Transformation. [International Journal of Engineering Research and Applications (IJERA), Vol. 3, Issue 3, May-Jun 2013].

A. Gebejes, R. Huertas analyzed the dependency of the features with the displacement considered for their computation and explored the possibility of features invariant under changes of the distance between the sample and observation position. Two experiments were performed – testing of the effect of the scale on the texture features and the definition of the best displacement value. Authors also suggested that the displacement giving the maximum Contrast (MCD) should be the best as it restricts the feature calculations to surely enclose only one texture element. [Conference of Informatics and Management Sciences (IJCTIC), March, 2013].

3. PROBLEM IDENTIFICATION

Keeping a focus on my research work, I found that face images fail to detect Lip Segmentation due to some Color Differences and Light Shadow near lip area and also eye feature extraction can match for two different persons sometimes.

4. METHODOLOGY OVERVIEW

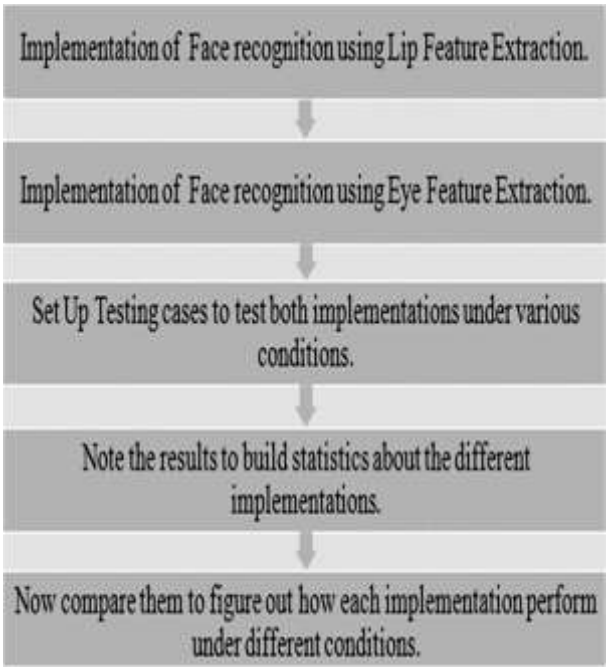


Figure 1. Methodology Overview of this paper

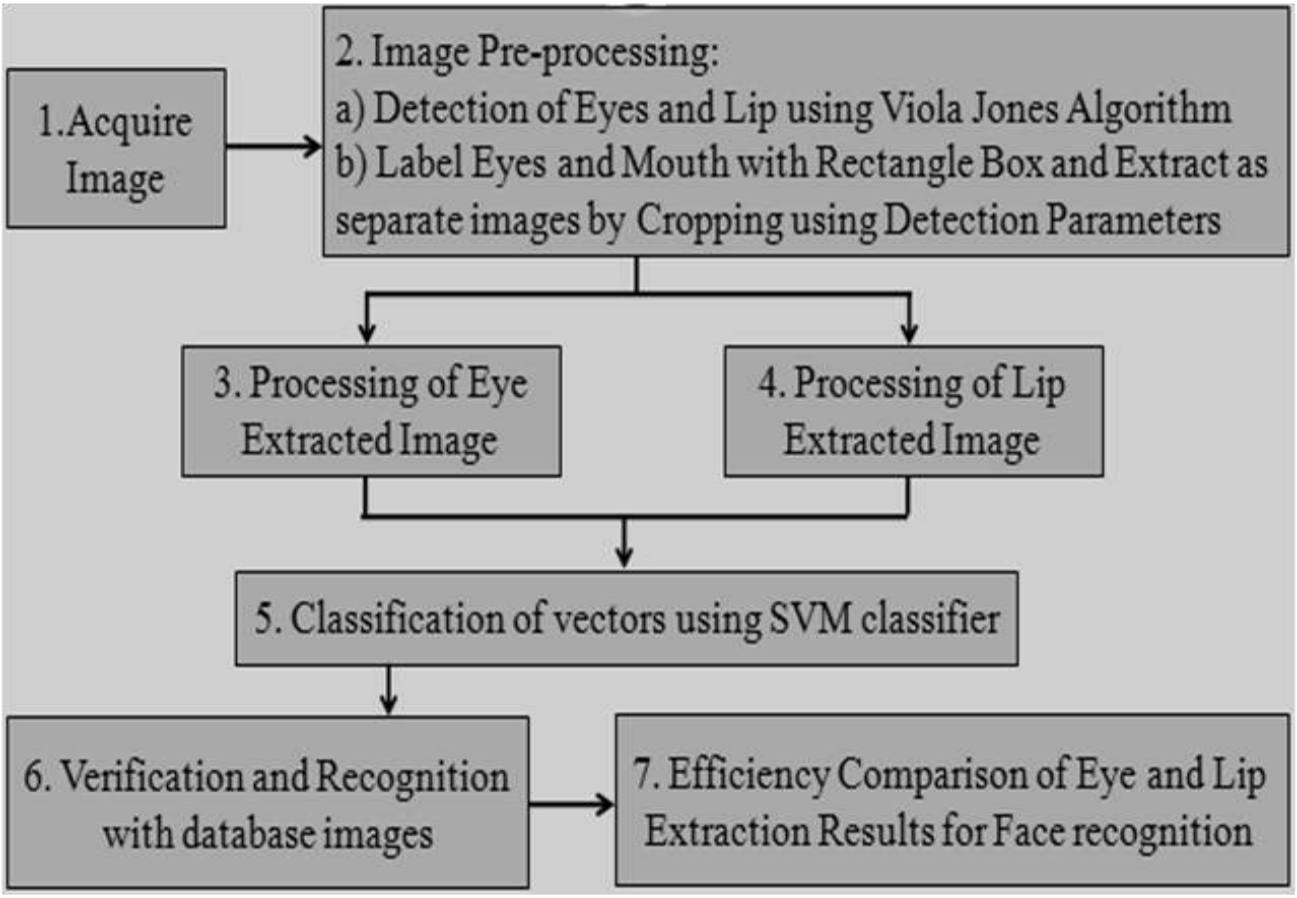


Figure 2. Proposed Methodology of this paper

5.1 Acquire Image

5.1.1 Input the image in MATLAB.
5.1.2 Scan the image in MATLAB for further processing.

5.2 Image Pre-processing

5.2.1 Detection of Eyes & Lip using Viola Jones Algorithm

5.2.1.1 Paul Viola and Michael Jone's presented a fast and robust method for face detection which is quicker than any technique with good accuracy. A widely used method for real-time object detection.
5.2.1.2 Training is slow, but detection is very fast.
5.2.1.3 Adaboost is used for object detection. Given a set of weak classifiers:-
None much better than random- Iteratively combine classifiers:

- Form a linear combination to get strong classifier.
- Training error converges to 0 quickly.
- Test error is related to training margin.

5.2.2 Label Eyes & Mouth with Rectangle box & Extract as separate images by Cropping using Detection Parameters

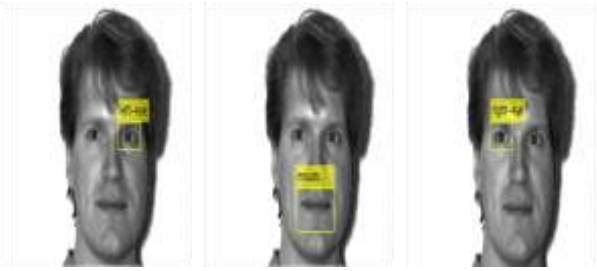


Figure 3. Lefteye, Mouth, Righteye labelled with rectangle box

Figure 4. Cropped Lefteye, Mouth, Righteye

5.3 Processing of Eye Extracted Image

5.3.1 Feature Extraction of Eyes using GLCM:

Gray Level Co-occurrence Matrix (GLCM) is created from a gray-scale image. Here the texture features of images are extracted and stored in a matrix. GLCM is one of the simplest matrix methods to extract the texture features. GLCM features are extracted for all the images in the database.
A co-occurrence matrix/ distribution, is defined over an image to be the distribution of co-occurring values at a given offset. The co-occurrence matrix is a statistical model and is useful in a variety of image analysis applications such as in biomedical, remote sensing, industrial defect detection systems, etc. Features computed from GLCM are based on the assumption that the texture information in an image is contained in the overall spatial relationship which grey levels of neighbouring pixels have to one another. GLCM contains information about the frequency of occurrence of two neighbouring pixel combination in an image. Although 22 features can be derived from GLCM, usually only some are considered as parameters of importance:

5.3.1.1 Contrast: It is a local grey level variation in the grey level co-occurrence matrix. It can be thought of as a linear dependency of grey levels of neighbouring pixels.

5.3.1.2 Dissimilarity: It is a measure that defines the variation of grey level pairs in an image. It is the closest to Contrast with a difference in the weight. Contrast unlike Dissimilarity grows 'Quadratically'.

5.3.1.3 Entropy: This in any system represents disorder, where in the case of texture analysis, it is a measure of its spatial disorder.

5.3.1.4 Energy: It is a measure of local homogeneity and therefore it represents the opposite of the Entropy. Basically this feature will tell us how uniform the texture is.

5.3.1.5 Prominence: It is a measure of asymmetry. When low, there is a peak in the GLCM matrix around the mean values.

5.3.1.6 Shade: It is a measure of the skewness of the matrix and is believed to follow the concepts of uniformity.

$$Contrast = \sum_{i,j} |i-j|^2 \, p(i,j)$$

$$Dissimilarity = \sum_{i,j} |i-j| \, p(i,j)$$

$$Entropy = -\sum_{i,j} p(i,j) \log\left(p(i,j)\right)$$

$$Energy = \sum_{i,j} p(i,j)^2$$

$$\mathrm{Pro} = \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} (i+j-u_x-u_y)^4 p(i,j)$$

$$\mathrm{Sha} = \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} (i+j-u_x-u_y)^3 p(i,j)$$

Figure 5. GLCM Features Equation

5.3.2 Feature Extraction of Eyes using ImageJ Toolbox:

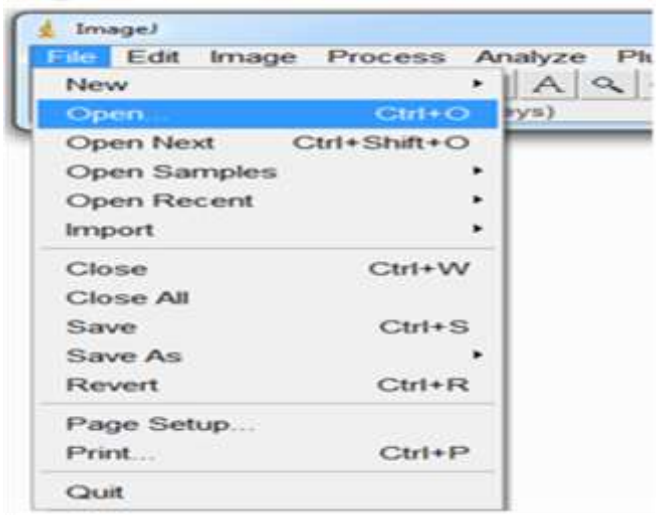


Figure 7. ImageJ Toolbox Screenshot 1

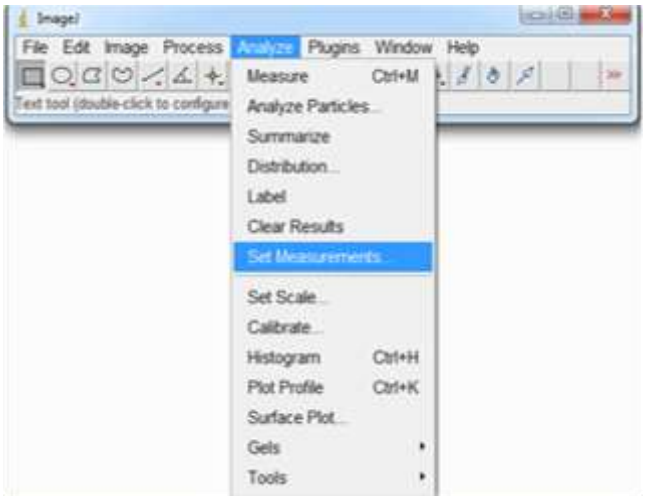


Figure 8.ImageJ Toolbox Screenshot 2

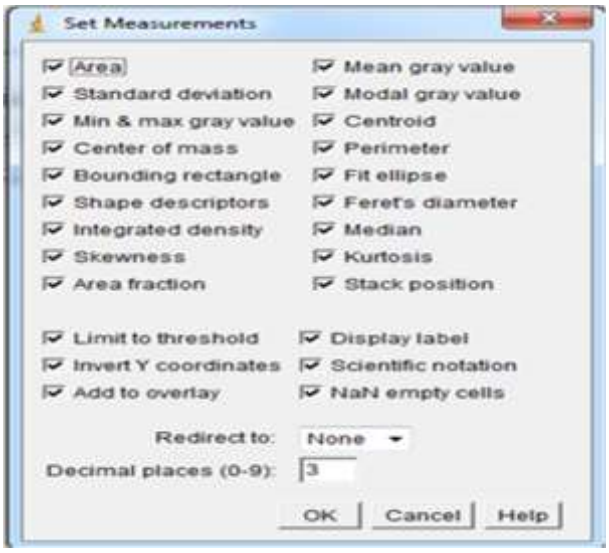


Figure 9. ImageJ Toolbox Screenshot 3

5.4 Processing of Lip Extracted Image

5.4.1 Feature Extraction of Lips using SIFT:

Scale Invariant Feature Transform (SIFT) is a method to detect local key points in a scene which are significantly important. SIFT key-points of objects are first extracted from a set of reference images and stored in a database.The technique has three steps:

5.4.1.1 Key-point detection by Difference of Gaussian (DoG) method:

We begin by detecting points of interest, which are termed key-points in the SIFT framework. The image is convolved with Gaussian filters at different scales, and then the difference of successive Gaussian blurred images is taken. Key-points are then taken as maxima/minima of the Difference of Gaussians (DoG) that occur at multiple scales. Once DoG images have been obtained, key-points are identified as local minima/maxima of the DoG images across scales. This is done by comparing each pixel in the DoG images to its neighbours at the same scale and corresponding neighbouring pixels in each of the neighbouring scales. If the pixel value is the maximum or minimum among all compared pixels, it is selected as a candidate key-point.

5.4.1.2 128 dimension key-point descriptor computation by analyzing orientation histogram:

Each key-point is assigned one or more orientations based on local image gradient directions. Previous step found key-point locations at particular scales and assigned orientations to them. This ensured invariance to image location, scale and rotation. Now we want to compute a descriptor vector for each key-point such that the descriptor is highly distinctive and partially invariant to the remaining variations such as illumination, 3D viewpoint, etc. This step is performed on the image closest in scale to the key-point's scale. Although the dimension of the descriptor, i.e. 128, seems high, descriptors with lower dimension than this don't perform as well across the range of matching tasks and the computational cost remains low due to the approximate method used for finding the nearest neighbour. Longer descriptors continue to do better but not by much and there is an additional danger of increased sensitivity to distortion and occlusion.

5.4.1.3 Key-point matching between two scenes by nearest neighbour method:

SIFT descriptors are invariant to minor affine changes. To test the distinctiveness of the SIFT descriptors, matching accuracy is also measured against varying number of key-points in the testing database, and it is shown that matching accuracy decreases only very slightly for very large database sizes, thus indicating that SIFT features are highly distinctive.

5.4.2 Applications of SIFT:

- Robot localization and mapping
- Panorama stitching
- 3D scene modelling, recognition and tracking
- 3D SIFT like descriptors for human action recognition
- Analyzing the Human Brain in 3D Magnetic Resonance Images

5.5 Classification of vectors using SVM classifier

A Support Vector Machine (SVM) is a classifier formally defined by a separating hyper-plane. In other words, given labeled training data (supervised learning), the algorithm outputs an optimal hyper-plane.

For a linearly separable set of 2D points which belong to one of two classes, find a separating straight line.

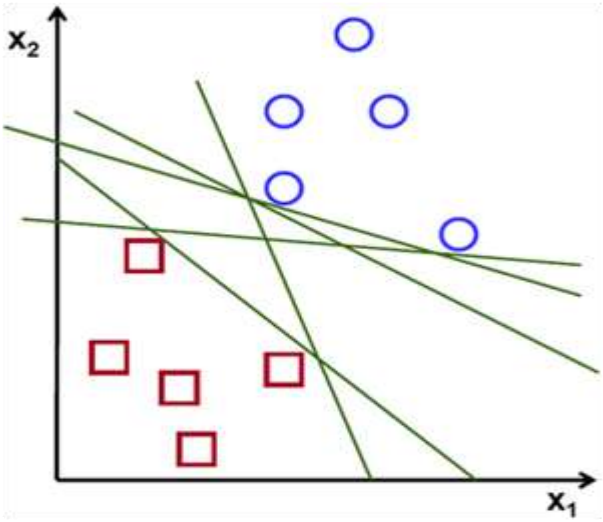


Figure 10. SVM - Separating Straight Line

A line is bad if it passes too close to the points because it will be noise sensitive and it will not generalize correctly. Therefore, our goal should be to find the line passing as far as possible from all points.

Then, the operation of the SVM algorithm is based on finding the hyper-plane that gives the largest minimum distance to the training examples.

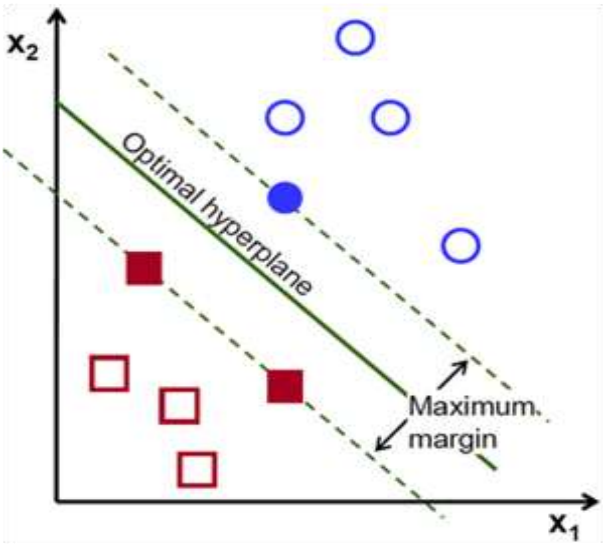


Figure 11. SVM - Optimal Separating Hyper-plane that Maximizes the Margin of the Training Data

5.5.1 ***SVM Algorithm:***

5.5.1.1 A hyper-plane gives a linear discriminant function in dimensions and splits the original space into two half-spaces:

$$h(x) = w^T x + b = w_1 x_1 + w_2 x_2 + \ldots + w_d x_d + b$$

Where,
w is a *d*-dimensional weight vector,
b is a scalar bias,
Points on the hyper-plane have $h(x) = 0$, i.e. the hyper-plane is defined by all points for which $w^T x = -b$

5.5.1.2 Given a separating hyper-plane h(x) = 0, it is possible to calculate the distance between each point x_i and the hyper-plane by:

$$\delta_i = \frac{y_i h(x_i)}{||w||}$$

5.5.1.3 The margin of the linear classifier is defined as the minimum distance of all points to the separating hyper-plane:

$$\delta^* = \min_{x_i} \{ \frac{y_i h(x_i)}{||w||} \}$$

All points (vectors x^*_i) that achieve this minimum distance are called the support vectors for the linear classifier. In other words, a support vector is a point that lies precisely on the margin of the classifying hyper-plane.

5.5.1.4 In a canonical representation of the hyper-plane, for each support vector x^*_i with label y^*_i we have:

$$y_i^* h(x_i^*) = 1$$

Similarly, for any point that is not a support vector, we have:

$$y_i h(x_i) > 1$$

By definition, it must be farther from the hyper-plane than a support vector. Therefore we have:

$$y_i h(x_i) \geq 1, \ \forall x_i \in D$$

The fundamental idea behind SVMs is to choose the hyper-plane with the maximum margin, i.e. the optimal canonical hyper-plane.
To do this, one need to find the weight vector *w* and the bias *b* that yields the maximum margin among all possible separating hyper-planes, that is, the hyper-plane that maximizes **1 / ||*w*||.** .
The problem then becomes that of solving a convex minimization problem (notice that instead of maximizing the margin, one can obtain an equivalent formulation of minimizing ||w||) with linear constraints, as follows:

Objective Function **Linear Constraints**

$$min \frac{||w||^2}{2} \qquad y_i h(x_i) \geq 1, \ \forall x_i \in D$$

This minimization problem can be solved using the *Lagrange multiplier* method, which introduces a Lagrange multiplier α for each constraint:

$$\alpha_i (y_i h(x) - 1) = 0 \text{ with } \alpha_i \geq 0$$

5.6 Efficiency Comparison of Eye and Lip Extraction Results for Face Recognition

Performance Evaluation Parameters - Comparison of:

5.6.1 Accuracy of Lip Extraction & Eye Extraction
5.6.2 Correct Matches in case of Lip Extraction & Eye Extraction
5.6.3 Wrong Matches in case of Lip Extraction & Eye Extraction

6. RESULTS

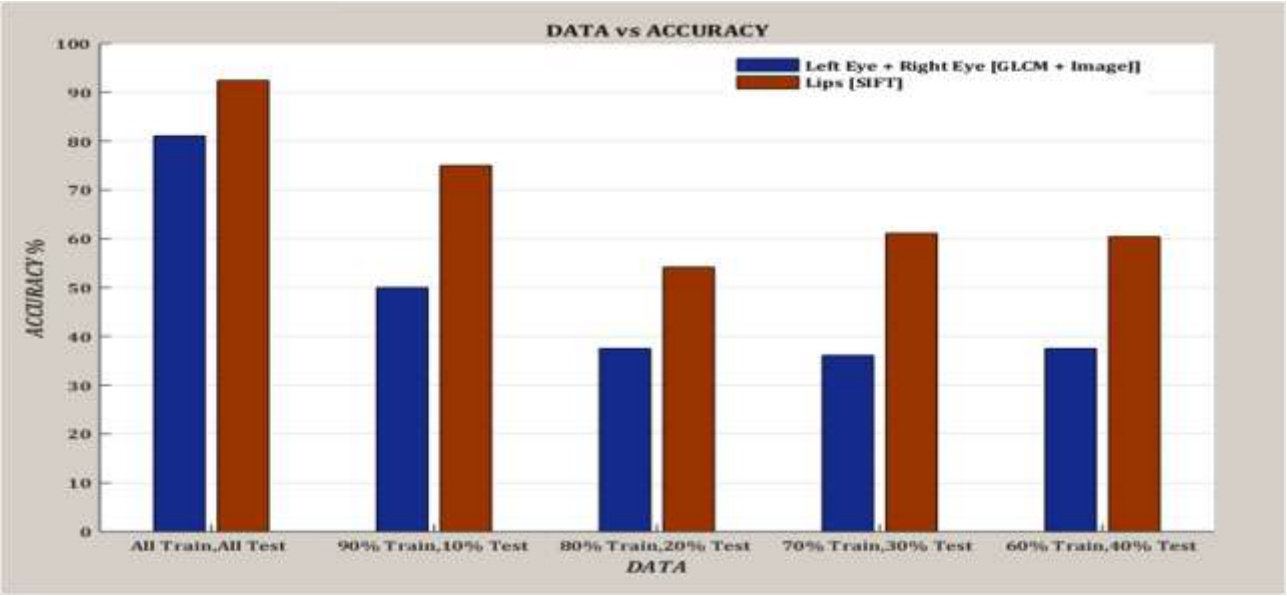


Figure 12. Data vs Accuracy

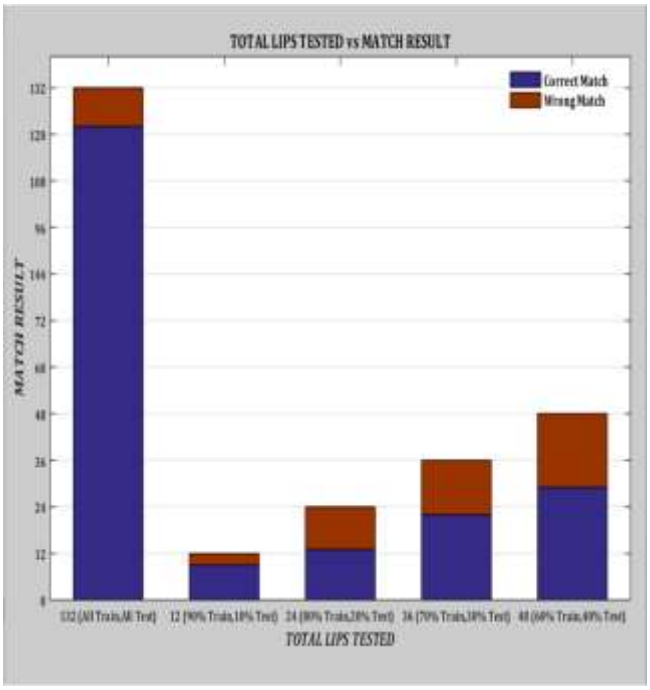


Figure 13. Mouth Performance

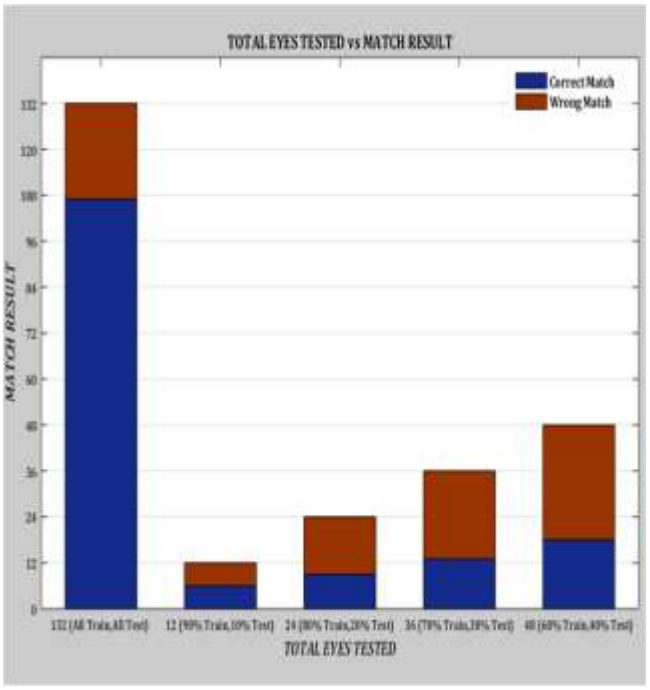


Figure 14. Eyes Performance

Table 1. Data vs Accuracy Plot Results

	LEFT EYE + RIGHT EYE METHOD [GLCM + IMAGEJ]			MOUTH METHOD [SIFT]			TOTAL DATASET	
DATA	**Accuracy**	**Correct**	**Wrong**	**Accuracy**	**Correct**	**Wrong**	**TRAIN**	**TEST**
All Train All Test	81.06%	107	25	92.42%	122	10	132	132
90% Train 10% Test	50%	6	6	75%	9	3	120	12
80% Train 20% Test	37.50%	9	15	54.17%	13	11	108	24
70% Train 30% Test	36.11%	13	23	61.11%	22	14	96	36
60% Train 40% Test	37.50%	18	30	60.42%	29	19	84	48

7. SCOPE OF FUTURE WORK

Face Recognition is a very vast and elaborated field. It has no end. As the advancement in science and technology, new techniques will continue developing day-by-day. Today, Lip and Eye extraction are mostly discussed techniques for face recognition purpose but in future many more advance techniques will arise for performing Face Recognition with much more accuracy and efficiency.

8. ACKNOWLEDGMENTS

My sincere thanks to all the respected and experienced faculties for their valuable guidance and motivation that always encouraged me to give my full dedication towards a new improvement in the field of science and image processing.

9. REFERENCES

[1] B. Sangeetha Devi and V. J. Arul Karthick, "Lip Recognition With Support Vector Machine (SVM) Classifier", International Journal of Advanced Research Trends in Engineering and Technology (IJARTET), Vol. II, Special Issue I, March 2015.

[2] Dr. S. Vijayarani, S. Priyatharsini, "Facial Feature Extraction Based On FPD and GLCM Algorithms", International Journal of Innovative Research in Computer and Communication Engineering, Vol.3, Issue 3, March 2015.

[3] Pankaj K.Sa, S.S. Barpanda, Bansidhar Manjhi, "Region Based Feature Extraction from Non-Cooperative Iris Images", Innovation Syst Software Engg, 2015.

[4] Sunil Sangve, Nilakshi Mule, "Lip Recognition for Authentication and Security", IOSR Journal of Computer Engineering (IOSR-JCE) Volume 16, Issue 3, Ver. VII, May-Jun. 2014.

[5] Rutuja G. Shelke, S. A. Annadate, "Face Recognition and Gender Classification Using Feature of Lips", International Journal of Innovation and Scientific Research (IJISR), Innovative Space of Scientific Research Journals (ISSR), Vol. 10, No. 2, Oct. 2014.

[6] Changcheng Li, Weidong Zhou, Shasha Yuan, "Iris Recognition Based on a Novel Variation of Local Binary Pattern", Springer-Verlag Berlin Hiedelberg, Vis Comput, 2014.

[7] Riddhi Patel and Shruti B. Yagnik, "A Literature Survey on Face Recognition Techniques", International Journal of Computer Trends and Technology (IJCTT), Volume 5, No.4, Nov 2013.

[8] Himanshu Srivastava, "Personal Identification Using Iris Recognition System", International Journal of Engineering Research and Applications (IJERA), Vol 3, Issue 3, May-Jun 2013.

[9] Ishvari S. Patel and Apurva A. Desai, "Lip Segmentation Based on Edge Detection Technique", International Journal of Scientific Research (IJSR), Volume II, Issue V, May 2013.

[10] Noopur Desai, Vidhi Dharia, Ishani Shah, Ghanshyam Prajapati, "Detecting the Location of Eyes on the Front Facial Images Using Two-Phase Method", International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE), Vol. 1, Issue 3, May 2013.

[11] A. Gebejes, R. Huertas, "Texture Characterization based on GLCM",Conference of Informatics and Management Sciences (IJCTIC), March, 2013.

[12] Jyoti Bedre and Shubhangi Sapkal, "Comparative Study of Face Recognition Techniques", Emerging Trends in Computer Science and Information Technology (ETCSIT2012), International Journal of Computer Applications (IJCA), 2012.

[13] Sasikumar Gurumurty and B. K. Tripathy, "Design and Implementation of Face Recognition System in Matlab Using the Features of Lips", I.J. Intelligent Systems and Applications (IJISA), July 2012 in MECS.

[14] Sambit Bakshi, Rahul Raman, Pankaj K Sa, "Lip pattern recognition based on local feature extraction", India Conference (INDICON), IEEE Annual, 2011.

[15] Zeynep Orman, Abdulkadir Battal, Erdem Kemer, "A Study on Face, Eye Detection and Gaze Estimation", International Journal of Computer Science & Engineering Survey (IJCSES) Vol.2, No.3, August 2011.

[16] Tanmay Rajpathak, Ratnesh Kumar, Eric Schwartz, "Eye Detection Using Morphological and Colour Image Processing", Florida Conference on Recent Advances in Robotics (FCRAR) 2009.

[17] Mayank Vatsa, Richa Singh, and Afzel Noore, "Improving Iris Recognition Performance Using Segmentation, Quality Enhancement, Match Score Fusion, and Indexing", IEEE Transactions on Systems, Man and Cybernetics— Part B: Cybernetics, Feb 2008.

[18] John Daugman, "New Methods in Iris Recognition", IEEE Transactions on Systems, Man and Cybernetics—Part B: Cybernetics, Vol. 37, No. 5, October 2007.

[19] Duy Nguyen, David Halupka, Parham Aarabi and Ali Sheikholeslami, "Real-Time Face Detection and Lip Feature Extraction Using Field-Programmable Gate Arrays", IEEE Transactions on Systems, Man and Cybernetics-Part B: Cybernetics, Vol. 36, No.4, August 2006.

[20] John Daugman, "How Iris Recognition works", IEEE Transactions on Circuits and Systems for Video Technology, Volume. 14, No. 1, Jan 2004.

[21] Mayank Vatsa, Richa Singh, Afzel Noore, "Improving Iris Recognition Performance Using Segmentation, Quality Enhancement, Match Score Fusion, and Indexing", IEEE Transactions on Systems, Man and Cybernetics—Part B: Cybernetics.

[22] Rafael C. Gonzalez, Richard E. Woods and Steven L. Eddins, "Digital Image Processing Using MATLAB", Second Edition

[23] The MathWorks, Inc., "Image Processing Toolbox", User's Guide, COPYRIGHT 1993–2015

[24] S. N. Sivanandam, S. Sumathi, S. N. Deepa "Neural Network using MATLAB 6.0".

[25] www.google.com

International Journal of Computer Applications Technology and Research
Volume 5– Issue 6, 338 - 346, 2016, ISSN:- 2319–8656

Recent Advances in Flower Pollination Algorithm

Nabil Diab
Department of Computer Science,
Faculty of Computers and Informatics,
Higher Technology institute 10th of Ramadan,
Egypt.

Emad El-Sharkawy
Faculty of Information Systems and Computer Science, October 6 University
Egypt.

Abstract: Flower Pollination Algorithm (FPA) is a nature inspired algorithm based on pollination process of plants. Recently, FPA has become a popular algorithm in the evolutionary computation field due to its superiority to many other algorithms. As a consequence, in this paper, FPA, its improvements, its hybridization and applications in many fields, such as operations research, engineering and computer science, are discussed and analyzed. Based on its applications in the field of optimization it was seemed that this algorithm has a better convergence speed compared to other algorithms. The survey investigates the difference between FPA versions as well as its applications. To add to this, several future improvements are suggested.

Keywords: flower pollination algorithm; inspired algorithms; optimization; hybridization

1. INTRODUCTION

Mathematical model for system modeling and objective function is used in most of the optimization algorithms but establishing mathematical model is not efficient due to its high cost of solution time, classical optimization algorithms are not efficient due to many causes such as ineffectiveness in adaptation of a solution algorithm, the dependency on type of objective and constraints, on type of variables used in modeling of problem, on solution space (convex, non-convex), number of decision variable, and the number of constraints in problem modeling that's why classical optimization algorithms are insufficient in large-scale combinational and non-linear problems. Most of algorithms solve models which have a certain type of object function or constraints. However, optimization problems in many different areas such as computer science, management, and engineering require concurrently different types of decision variables, object function, and constraints in their formulation. Therefore, meta-heuristic optimization algorithms become quite popular methods in recent years, because they provide good computing power and easy conversion, and due to its flexibility to convert meta-heuristic program from a single objective function problem to a multi objective problem or a different problem.

Recently, Yang [1] developed a new Flower pollination algorithm that draws its inspiration from the flow pollination process of flowering plants. FPA testing results in many fields proved their ability to be used in a wide range of optimization problems and also their ability to provide better performance in comparison with other traditional optimization techniques. This paper introduces improvements, hybridization and applications of FPA.

This paper is organized as follows: after introduction, the original Flower pollination algorithm is briefly introduced. In section 3, Multi-objective FPA is described. Section 4 the Binary FPA version is introduced. The algorithm improvements are taken in section 5, while the real-world applications of FPA are discussed in section 6. Finally, conclusions and suggestions for future work are presented in section 7.

2. FLOWER POLLINATION ALGORITHM

FPA, inspired by the flow pollination process of flowery plants, was developed in 2012 by Xin-She-Yang [1]. The following 4-rules are used as a matter of convenience.

1.Biotic and cross-pollination is considered as global pollination process with pollen-carrying pollinators performing L´evy flights.

2. Abiotic and self-pollination are considered as local pollination.

3. Flower constancy can be considered as the reproduction probability is proportional to the similarity of two flowers involved.

4. Local pollination and global pollination is controlled by a switch probability p ∈ [0, 1].

In order to formulate updating formulas, we have to convert the aforementioned rules into updating equations. For example, in the global pollination step, flower pollen gametes are carried by pollinators such as insects, and pollen can travel over a long distance because insects can often fly and move in a much longer range[1].Therefore, Rule 1 and flower constancy can be represented mathematically as:

$$x_i^{t+1} = x_i^t + \gamma L(\lambda)(x_i^t - B) \quad (1)$$

Where x_i^t is the pollen i or solution vector xi at iteration t, and B is the current best solution found among all solutions at the current generation/iteration. Here γ is a scaling factor to control the step size. In addition, L(λ) is the parameter that corresponds to the strength of the pollination, which essentially is also the step size. Since insects may move over a long distance with various distance steps, we can use a Le'vy flight to imitate this characteristic efficiently. That is, we draw L > 0 from a Levy distribution:

$$L \sim \frac{\lambda\Gamma(\lambda)\sin(\pi\lambda/2)}{\pi}\frac{1}{S^{1+\lambda}}, (S >> S_0 > 0) \quad (2)$$

Here, Γ(λ) is the standard gamma function, and this distribution is valid for large steps s > 0.

Then, to model the local pollination, both Rule 2 and Rule 3 can be represented as:

$$x_i^{t+1} = x_i^t + U(x_j^t - x_k^t) \quad (3)$$

Where x_j^t and x_k^t are pollen from different flowers of the same plant species. This essentially imitates the flower constancy in a limited neighborhood. Mathematically, if x_j^t and x_k^t comes from the same species or selected from the same population, this equivalently becomes a local random walk if we draw U from a uniform distribution in [0, 1].Though Flower pollination activities can occur at all scales, both local and global, adjacent flower patches or flowers in the not-so-far-away neighborhood are more likely to be pollinated by local flower pollen than those faraway. In order to imitate this, we can effectively use the switch probability like in Rule 4 or the proximity probability p to switch between common global pollination to intensive local pollination. To begin with, we can use a naive value of p = 0.5as an initially value. A preliminary parametric showed that p = 0.8 might work better for most application. The pseudo code of FPA is shown in Figure 1.

Flower Pollination Algorithm
Objective function min or max f(x), x=(x1, x2, ..., xd)
Initialize a population of n flower/pollen gametes with random solutions.
Find the best solution in the population
Define a switch probability
Define a stopping criterion
***While** ()*
***For** i=1 : n (all n flowers in the population)*
***If** rand<p,*
Draw a (d-dimensional) step vector L which obeys Levy distribution
Do Global pollination via
Else
Draw from a uniform distribution in [0,1]
Do Local pollination via
End If
Evaluate new solutions
If new solutions are better, than update them in the population
Find current best solution
End While

Fig.1: Pseudo code of standard Flower Pollination Algorithm (FPA).

To simplify the implementation of this algorithm it is assumed that each flower produces only one flower ,There are some applications that require to assign each flower with multiple pollen gametes and each plant with multiple flowers such as Graph coloring , image compression and multi-objective optimization problems.

3. FPA MULTI-OBJECTIVE VERSION [3]

Multi-objective optimization problems are typically complex problems. The methods for solving multi-objective problems differ from algorithms for single objective optimization.
For single objective optimization, the optimal solution can often be a single point in the solution space, while for bi-objective optimization, the Pareto front forms a curve, and for tri-objective cases, it becomes a surface. In fact, higher dimensional problems can have extremely complex hyper-surface as its Pareto front ([4];[5];[6];[7]) Consequently, it is typically more challenging to solve such high-dimensional problems. Real-world design problems in engineering and industry is considered as multi-objective optimization that requires to produce many points on the Pareto front for good approximations. In order to compare the performance of the proposed MOFPA with other multi-objective algorithms, we have selected a few algorithms with available results from the literature, such as vector evaluated genetic algorithm (VEGA) [8],NSGAII[9],multi-objective differential evolution (MODE)([10];[11]),differential evolution for multi-objective optimization(DEMO)[12],multi-objective bees algorithms[13], and strength Pareto evolutionary algorithm(SPEA) [14].
The experimental results for 11 test functions and two design examples suggest that MOFPA is a very efficient algorithm for multi-objective optimization. The algorithm can deal with highly nonlinear, multi-objective optimization problems with complex constraints and diverse Pareto optimal sets.

4. FPA BINARY VERSION

Binary-constrained version of the Flower Pollination Algorithm [15] (FPA) is implemented for feature selection, BFPA determines the features that compose the final set by using Boolean search space, this method was applied to some public and private datasets. Numerical experiments in comparison with Particle Swarm Optimization, Harmony Search and Firefly Algorithm have demonstrated the suitability of the FPA for feature selection.

5. IMPROVEMENTS

There are a lot of significant research studies that aim to improve the FPA performance, by algorithm modification or effective hybridization with other Nature Inspired Algorithms. The improved algorithms support researchers to solve more optimization problems.

5.1 Modification

5.1.1 Modified flower pollination algorithm (MFPA) [16]

After testing the FPA they found an important impact on the computational burden and convergence solution in the initial population and switching from local to global pollination.
The following modifications improve the algorithm performance:

A. Looking for the Best Initial Condition
B. Switching the Local to Global Pollination Process

This modification eliminates the use of the probability switch and combines Lévy flights with Brownian motion into a single random walk equation. MFPA has been successfully implemented to solve optimal power flow problem. The algorithm has been tested on the IEEE 30-bus system. The experimental results show results enhancement compare to other meta-heuristics algorithms. The main advantage of the MFPA is to find fitter initial solutions with improvement of the switching process. Both modifications let the algorithm reach the convergence with minimum number of iterations and little processing time.

5.1.2 Improved Flower Pollination Algorithm with Chaos [17]

The definite integrals are useful in a wide range of applications in many regions such as operation research, computer science, mathematics, mechanics, physics, and engineering. Improved Flower pollination algorithm with chaos (IFPCH) introduced by [17] for solving definite integrals. Numerical results show that the algorithm offers an effective way to calculate numerical value of definite integrals, it can converge to the best solution and it has a high convergence rate, high accuracy and robustness.

5.2 Hybridization

After making several hybridization between the flower pollination algorithm and other inspired meta-heuristic algorithms, there is certainty that Hybridization improve the performance of the FPA to reach the optimal solution faster than the algorithm itself in a short period of time and to reduce the algorithm limitations.

5.2.1 Flower pollination algorithm with Chaotic Harmony Search (FPCHS) [18]

Integration of flower pollination algorithm with chaotic harmony search aims to improve the searching accuracy, the planned algorithm is used to solve Sudoku puzzles. The algorithm has been tested on a set of Sudoku problems. The results verify that the planned algorithm is more efficient compared with other algorithms.

5.2.2 Flower pollination algorithm with genetic algorithm (FPA-GA) [19]

In the FPA-GA the flower pollination algorithm (FPA) and genetic algorithm (GA) are integrated to solve constrained optimization problems. The experimental results of testing the FPA-GA in seven benchmark optimization problems verify that the proposed algorithm is very effective and can be extended to solve other problem such as combinatorial optimization problems.

5.2.3 Flower pollination algorithm with particle swarm optimization (FPPSO) [20]

Integration of Flower pollination algorithm with particle swarm optimization aims to improve the searching accuracy, the hybrid algorithm is used to solve constrained optimization problems, FPPOS has been validated using several benchmark mathematical and engineering design problems and the results prove that the FPPOS is more efficient at finding global optimal solution than other algorithms.

5.2.4 Flower pollination algorithm with differential evolution optimization algorithm (DE-FPA) [21]

New hybrid population based algorithm which has of the exploration capability of differential evolution and also inherits the directed movement towards the global best and local search from the flower pollination algorithm this proposed algorithm was handling the drawbacks of the individual algorithms such as slow convergence to the global optima at the differential evolution .DE-FPA has been tested with some benchmark test functions ,the results clearly establish that the DE-FPA have the superiority over both the DE and FPA in terms of performance and convergence rate. And a novel concept of dynamic adaptive weight is introduced for faster convergence than the individual algorithms.

5.2.5 Flower pollination algorithm with K-Means algorithm (FPAKM) [22]

FPAKM is a hybrid data clustering approach using Flower Pollination Algorithm and K-Means, K-Means algorithm solve the data clustering problems in a very fast way but find the local optimum solution , the flower pollination algorithm is the global optimization technique that eliminate this drawback , FPAKM has been tested on eight datasets , The experimental results prove that FPAKM find optimal cluster centers and that is better than these individual algorithms.

5.2.6 Flower pollination algorithm, differential evolution algorithm (HFPA) with time varying fuzzy selection mechanism (HFPA-TVFSM) [23]

In the HFPA the flower pollination algorithm (FPA) and differential evolution (DE) algorithm are integrated to solve wind-thermal dynamic multi-objective optimal dispatch problem that improves the exploration and exploitation potential of the flower population which is conducting the search, in this problem there are more constraint such as minimization of cost, emission and losses , and other complex constraints like valve point loadings, ramp limits, prohibited zones and spinning reserve . To solving multi-objective problems HFPA is integrated with a 5-class, 3-step time varying fuzzy selection mechanism (TVFSM) that effectively searches the best compromise solution (BCS) .HFPA-TVFSM is tested and validated on two wind-thermal test systems from literature , the results prove that the hybrid algorithm performs efficiently and all constraints are satisfied.

5.2.7 Flower pollination algorithm by Artificial Bees (FPAB) and Biogeography Based Optimization Algorithm (FPAB/BBO) [24]

Remote sensing has been used for the classification of satellite image on a very large scale. A hybridization of Flower Pollination by Artificial Bees (FPAB) and Biogeography Based Optimization Algorithm has been used to classify a satellite image as an efficient land cover classifier for satellite image ,this resulted in highly accurate classification of the satellite image is obtained by using FPAB/BBO and many of the land cover features are identified much more clearly

5.2.8 Flower pollination algorithm with Clonal Selection Algorithm (MFPA) [25]

The Flower Pollination Algorithm is hybridized with the Clonal Selection Algorithm, The Experimental results of testing the Algorithm in 23 optimization benchmark problems verify the algorithm superiority compared to other famous optimization algorithm such as Simulated Annealing, Genetic Algorithm, Flower Pollination Algorithm, Bat Algorithm, and

Firefly Algorithm , and also verify that the algorithm is able to find more accurate solutions than the other optimization techniques.

5.2.9 Flower Pollination Algorithm with Simulated Annealing Algorithm (FPSA) [26]

The standard flower pollination algorithm (FPA) with simulated annealing are integrated to enhance the search performance and speeds up the global convergence rate, FPA generate initial state of simulated annealing and new solutions , FPSA has been tested in structural engineering optimization problems. The experimental results prove that this method is accurate in finding the best solution and have fast convergence speed compared to existing algorithms.

5.2.10 Flower Pollination Algorithm with Tabu Search Algorithm (TS-FPA) [27]

Integration of flower pollination algorithm (FPA) and Tabu Search Algorithm to solve unconstrained optimization problems , these technique eliminate the limitations of FPA in population diversity and convergence precision . TS-FPA is validated by ten benchmark functions. The experimental results show the algorithm high stability and it also has a fast convergence speed.

6. APPLICATIONS

FPA is validated in several applications and all the experimental results are more efficient compared to other algorithms ,Furthermore FPA finds global optimal solution in a short period of time. In recent years ,FPA has been applied by several researchers to various optimization problems in computer science, operations research , engineering and science among which we mention the important applications.

6.1 Computer Science

Computer science is the scientific and practical approach to computation, computer science aims to identifying certain computer science concepts that can be used directly in solving real world problems. Meta-heuristic algorithms are a higher level procedure or heuristic designed to find, generate, or select a heuristic. Metaheuristic algorithms applied in many CS applications provide a sufficiently good solution to optimization problems.
FPA is a nature inspired metaheuristic algorithm implemented in CS applications to find optimal solution .The numerical results clearly show that the FPA gives better results that prove the algorithm suitability for solving different real world problems .
Some applications of FPA in Computer science field:

6.1.1 Robust and Efficient ‘RGB’ based Fractal Image Compression [28]

Pollination Based Optimization (PBO) is a new technique based on Pollination Based Optimization which improve the fractal image compression FIC ,these technique used to classify the phantom, satellite and rural image dataset .the experimental results show that PBO performance is better than other optimization technique ,using PBO reduce the encoding time and improves the visual quality of image in comparison with FCI.PBO results encouraged researchers to use these technique many area of the multimedia system’s world.

6.1.2 Wireless Sensor Network Lifetime Global Optimization [29]

To achieve the global optimization for WSN lifetime using flower pollination algorithm rather than the classical LEACH algorithm .FPA selects the best CHs distribution that guarantees a routing optimization with the minimum communication links` cost between nodes within each cluster. This also optimize the sum of nodes energy (WSN total energy) based on the fixed number of cluster heads that minimize the overhead of data aggregation and the management and control signals for each cluster head. Results show an enhancement in the wireless sensor network lifetime ,network stability and overall network lifetime.

6.1.3 Retinal vessel segmentation [30]

FPA is implemented to search for the optimal clustering of the given retinal image into compact clusters under some constrains ,at local search period. Shape features are used to enhance the clustering results, this approach is tested on a publicly available databases drive of retinal images. Results prove that performance enhancement compared to other art techniques.

6.1.4 Graph coloring problem [31]

Flower pollination algorithm is a recent nature inspired algorithm for continuous optimization, a discrete version of the algorithm was used to solve graph coloring problem. The performance of the algorithm is evaluated on benchmark instances set, the experimental results ensure it achieves the exact solution in almost cases.

6.1.5 Planar Graph Coloring Problem [31,32]

Local greedy flower pollination algorithm (LGFPA) was implemented for solving planar graph coloring problem based on local swap strategy, local reverse strategy. These algorithm can solve the planar graph coloring problem using four-colors more efficiently and accurately. The experimental results show that this algorithm can get smaller average iterations than basic flower pollination algorithm (FPA), particle swarm optimization (PSO), differential evolution (DE), and can obtain higher successful coloring rate.

6.2 Operation research

Operation research is often considered to be a sub-field of mathematic aim to find optimal or near-optimal solutions to complex decision-making problems by using optimization techniques .Many meta-heuristic optimization methods have been developed to solve such difficult optimization problems. FPA is a nature inspired metaheuristic algorithm implemented in OR applications to solve optimization problems.
The numerical results clearly show that the FPA gives better results that prove the algorithm suitability for solving different difficult optimization problems.
Some applications of FPA in Operation research field:

6.2.1 Integer Programming Problems [33]

Integration of the flower pollination algorithm with chaos theory (IFPCH) to solve integer programming problems, the new solution produce by FPA taken by the IFPCH that rounded the parameter values to the closest integer .This approach has been tested by several problems, the experimental results prove the superiority of the new approach over the standard FPA in terms of efficiency and success rate.

6.2.2 Large Integer Programming Problems [33, 34]

An improved version of Flower pollination Meta-heuristic Algorithm, (FPPSO) that combines the standard flower pollination algorithm (FPA) with the particle swarm optimization (PSO) algorithm to improve the searching accuracy, using this algorithm to solving integer programming problems result in obtain the optimal solution within less computation.

6.2.3 Ratios Optimization Problems [35]

Integration of the Flower Pollination Algorithm with Chaos Theory (IFPCH) to solve large-scale ROP with an optimal solution at a finite point and an unbounded constraint set .The technique is tested using several ROP benchmark. The test aims to prove the capability of the IFPCH to solve any type of ROPs. The experimental results show the feasibility, effectiveness, and robustness of the technique. The results revealed the superiority of the technique among others in computational time.

6.2.4 Dimension by Dimension Improvement [36]

FPA has been work with three optimization strategies (local neighborhood search strategy, dimension by dimension evaluation and improvement strategy, and dynamic switching probability strategy) to improve FPA deficiencies .The experimental results by applied 12 typical standard benchmark functions simulation show that DDIFPA algorithm has strong global searching ability and local optimization ability, also can improve the convergence speed and the quality of solutions effectively.

.6.2.5 Optimal Unmanned Undersea Vehicle Path Planning Problem [37]

Implementation of flower pollination algorithm to solve the UUV path planning problem in 2D and 3D space. Optimization strategies of particle swarm optimization are applied to the local search process of IFPA to enhance its search ability, and the solution exposed to dimension by dimension based update and evaluation strategy .This approach can accelerate the global convergence speed. The experiment results prove that this approach is more effective and feasible in UUV path planning in 2D and 3D space compared to nine population based algorithms.

6.2.6 Selection of University academic credits [38]

Implementation of FPA to make optimal selection of an elective in Bachelor level and to interact with the different values associated with the achievement of the term loans and the cost-benefit for every student in a minority group and comparing their choices of electives with respect to the group. The algorithm is tested in case of four scholar minorities studies in a University with approximately 87 educational on Bachelor level.

6.3 Electrical Engineering

In electric power market, it is too difficult to offer electric power to customer with high quality and least cost because of several complicated problems. There is a need to find optimal solution for operation and design of an efficient power system. To find optimal solution FPA is implemented for different problems. The numerical results clearly show that the FPA gives better results that prove the algorithm suitability for different power system optimization problems.
Some applications of FPA in electrical field:

6.3.1 Optimal Reactive Power Dispatch Problem [39]

Optimal Reactive Power Dispatch Problem ORPD objective is reduce losses and stability of the system improvement. FPA is implemented in the standard IEEE-30 bus system and results are compared to other recently reported work and it is clear that this algorithm performs in a better way.

6.3.2 Power Loss Reduction on Radial Distribution System problem [40]

FPA implemented to decides the locations and size of capacitors to realize the optimum sizable reduction in active power loss and significant improvement in voltage profile. This method is tested on 10, 15, 69 and85-bus radial distribution systems, the results found is better than the other methods in terms of the quality of solution.

6.3.3 Optimal Line Flow in Conventional Power System problem [41]

Euclidean affine flower pollination algorithm (EFPA) used to addresses the optimal line flow OLF constraint for minimizing the fuel cost, loss, emission and voltage stability index for all generators in distributed power systems ,EFPA is performed on IEEE 30 bus system and IEEE 57 bus system, Results proved that the EFPA optimization is efficient for OLF constraint.

6.3.4 Linear antenna array optimization problem [42]

Apply FPA to the electromagnetic and antenna community for the optimization of linear antenna arrays. FPA was applied to obtain optimized antenna positions in order to achieve desired array pattern with minimum SLL along with null placement in specified directions. Results have been compared to conventional array (non-optimized), and with arrays optimized using other nature-inspired evolutionary algorithms such as ACO, CSO and PSO, Consequent FPA outperforms the other evolutionary algorithms and at times it yields a similar performance.

6.3.5 Optimal Control in Multi-Machine System with GUPFC Problem [43]

Generalized unified power flow controller (GUPFC) problem is to tuning the control parameters of PSS and POD, Flower pollination algorithm implemented in multi-machine system to provide parameters of two stage lead-lag controller .The results demonstrate that Flower pollination algorithm can be used as a method of tuning the control parameters of PSS and POD in a multi-machine power system.

6.3.6 Synthesis of Circular Array Antenna for Side lobe Level and Aperture Size Control [44]

FPA is implemented for circular array synthesis to solve Side lobe level suppression, FPA adapts one and two degrees of freedom, namely, amplitude only and amplitude spacing. The experimental results prove the effectiveness of FPA compared to genetic algorithm (GA) and uniform circular array antenna (UCAA)

6.3.7 Improvement of Voltage Profile using Shunt Capacitor in Radial Distribution System [45]

A method of reducing the loss by placing shunt capacitors at the optimal location, using Loss Sensitivity Factor and FPA to determine capacitor placement make significant decrease in power loss, increase in voltage profile and decrease in total annual cost.

6.3.8 Voltage Profile Improvement using Distributed Generation in 33-bus and 69-bus RDS [46]

A method of reducing the loss by placing distribution generation at the optimal location , using Loss Sensitivity Factor and FPA to determine optimal allocation of distribution generation in a radial distribution system make significant decrease in power loss, increase in voltage profile ,this method is tested on 33-bus and 69-bus system, yields good results.

6.3.9 Optimal sizing and locations of capacitors in radial distribution systems [47]

Implementation of Flower Pollination Algorithm to allocate and determine capacitors sizes in various distribution systems. FPA is tested on 15, 69 and 118-bus radial distribution systems ,results is better than the other algorithms such as Genetic Algorithm (GA), Particle Swarm Optimization (PSO),Plant Growth Simulation Algorithm (PGSA), Direct Search Algorithm (DSA), Teaching Learning-Based Optimization(TLBO), Cuckoo Search Algorithm (CSA), Artificial Bee Colony (ABC) and Harmony Search. FPA enhance the voltage profile , minimize the losses and total cost for various distribution systems.

6.3.10 optimal placement and sizing of distributed generation in Distribution systems [48]

A method of reducing the loss by placing Distributed generator at the optimal location and determine the optimal DG size, using the index vector method and FPA to determine optimal allocation of Distributed generator in a radial distribution system and determine the optimal DG size make significant decrease in power loss, increase in voltage profile, this method is tested on 15-bus 34-bus and 69-bus system, yields good results.

6.3.11 Distributed generations planning [49]

A method of reducing the loss by placing Distributed generator at the optimal location and determine the optimal DG size, using the index vector method and FPA to determine optimal allocation of Distributed generator in a radial distribution system and determine the optimal DG size make significant decrease in power loss, increase in voltage profile, this method is tested on 33-bus 69-bus and 136-bus system, yields good results. Furthermore, the results obtained by the FPA algorithm are compared with other metaheuristic optimization techniques such as backtracking search optimization algorithm, artificial bee colony, and selection algorithm ,The outcomes verify that the FPA algorithm is efficient.

6.3.12 Load Frequency Control for a Hydro-Thermal Deregulated Power System [50]

FPA is implemented to tune the PI controller of the LFC to improve the dynamic response, by minimizing the Integral Square Error of the system. The results show that the FPA tuned PI controller improves the dynamic response of the deregulated system faster than the PI controller for different cases.

6.3.13 Economic and emission dispatch problems [51]

Implementation of Flower Pollination Algorithm (FPA) to solve ELD and CEED problems in power systems, Results prove that FPA have the superiority over other optimization algorithms for large scale power system with valve point effect in terms of total cost and computational time.

6.3.14 Different Economic Load Dispatch Problems [52]

Flower pollination algorithm is implemented to adjust real power generations for minimizing the fuel cost, FPA is tested on the standard IEEE-30 bus system and the results are compared with those of the other algorithms, The results are found improved and encouraging.

6.3.15 Short Term Hydrothermal Scheduling [53]

Implementation of improved flower pollination algorithm (IFPA) for solution of short term hydrothermal scheduling problem. In IFPA the local pollination process of FPA is controlled by adding a scaling factor and an additional intensive exploitation phase is added to tune and improve the best solution. The improved algorithm and FPA are tested by three different test cases .the experiment result provide the superiority of IFPA.

6.4 Other Fields

Applications of FPA in other fields;

6.4.1 Estimation of the Transition Matrix in Markov Chain Model of Customer Lifetime Value [54]

FPA is implemented early in application about real health insurance data and it give efficient results in solving inverse problem in Markov Chain Model. Consequently FPA is applied to solve inverse problem of Customer Lifetime Value (CLV). For the study case, health insurance data is taken along with some arbitrary constant interest rates for the next five years.

6.4.2 Application of the Flower Pollination Algorithm in Structural Engineering [55]

Structural engineering problems have various design constraints concerning structural security measures and practicability in production. Thus, optimization becomes an important part of the design process. FPA was implemented to solve structural engineering problems. The engineering problems are about pin-jointed plane frames, truss systems, deflection minimization of I-beams, tubular columns, and cantilever beams. The experimental results show that FPA is effective to find the best optimum results when compared to other method.

7. CONCLUSION AND FUTURE WORK

How to solve real world optimization problems is a question asked every day and researchers seek to find the answer ,therefore implementing Nature inspired algorithm has being one of the effective methods to solve these problems, since 2012 the Flower Pollination Algorithm proposed by Xin-She Yang [1] was implemented in many fields such as operation researches ,computer science and electrical engineering. This paper is a review which explores the FPA efficiency whether by implementation or by hybridization with other algorithms, the studies prove that FPA is a powerful tool in solving several optimization problems, a consequence FPA is one of the more commonly used algorithms until now. The future work will be focused on:

i. Applying binary flower pollination version to solve most popular constrained engineering optimization problems.

ii. Applying binary flower pollination version to solve combinatorial optimization problem such as (Quadratic assignment problem, Traveling salesman problem, Graph Coloring problem, Knapsack problem, etc.).

iii. Applying FPA with fuzzy logic.

iv. Combining FPA with Neutrosophic logic.

v. Applying FPA to solve geographical information system applications

vi. Applying FPA to solve machine learning problems.

vii. Proposing new adaptive mechanism to update FPA parameters.

viii. Analyzing the effect of hybridizing FPA with other EAs.

REFERENCES

[1] Xin-She Yang, “Flower pollination algorithm for global optimization”, in: Unconventional Computation and Natural Computation 2012, Lecture Notes in Computer Science, Vol. 7445, pp. 240-249 (2012).

[2] Wikipedia article on pollination, http://en.wikipedia.org/wiki/Pollination.

[3] X. S. Yang, M. Karamanoglu, X. S. He,” Flower Pollination Algorithm: A
Novel Approach for Multiobjective Optimization” in: Engineering Optimization, vol. 46, Issue 9, pp.1222-1237 (2014).

[4] Madavan, N. K., 2002. “ Multiobjective optimization using a pareto differential evolution approach.” in: Congress on Evolutionary Computation (CEC'2002), Vol. 2, NewJersey, IEEE Service Center, pp. 1145{1150}

[5] Marler, R. T. and Arora, J. S., 2004. “ Survey of multi-objective optimization methods for engineering.” In: Struct. Multidisc. Optimal, 26:369{395}

[6] Yang, X. S., 2010a. Engineering Optimization: An Introduction with Metaheuristic Applications, John Wiley and Sons, USA.

[7] Yang, X. S. and Gandomi, A. H., 2012. “Bat algorithm: a novel approach for global engineering optimization.” Engineering Computations, 29(5):464{483}

[8] Schafier, J. D., 1985. “Multiple objective optimization with vector evaluated geneticalgorithms.” in: Proc. 1st Int. Conf. Genetic Algorithms, pp. 93{100}

[9] Deb, K., Pratap, A., and Moitra, S., 2000. “Mechanical component design for multiple objectives using elitist non-dominated sorting GA”, in: Proceedings of the ParallelProblem Solving from Nature VI Conference, Paris, 16-20 Sept 2000, pp. 859-868.

[10] Babu, B. V. and Gujarathi, A. M., 2007. “Multi-objective differential evolution (MODE) for optimization of supply chain planning and management”, in: IEEE Congress on Evolutionary Computation (CEC 2007), pp. 2732-2739.

[11] Xue, F., 2004.” Multi-objective differential evolution: theory and applications”, PhD the-sis, Rensselaer Polytechnic Institute.

[12] Robific, T. and Filipific, B., 2005. “ DEMO: differential evolution for multi-objective optimization.” in: EMO 2005 (eds. C. A. CoelloCoello), LNCS 3410:520{533}

[13] Pham, D. T. and Ghanbarzadeh, A., 2007.” Multi-Objective Optimization using the Bees Algorithm.” in: 3rd International Virtual Conference on Intelligent Production Machines and Systems (IPROMS 2007), Whittles, Dunbeath, Scotland.

[14] Deb, K., Pratap, A., Agarwal, S., Mayarivan, T., 2002.” A fast and elististmul-tiobjective algorithm: NSGA-II”, IEEE Trans. Evol. Computation, 6(2):182-197.

[15] Douglas Rodrigues, Xin-She Yang, André Nunes de Souza and João Paulo Papa,” Binary Flower Pollination Algorithm and Its Application to Feature Selection”, Recent Advances in Swarm Intelligence and Evolutionary Computation, Studies in Computational Intelligence 585, DOI 10.1007/978-3-319-13826-8_5

[16] José A. Regalado, Emilio Barocio E, Erick Cuevas , “Optimal Power Flow Solution Using Modified Flower Pollination Algorithm” Conference Paper · November 2015 ,978-1-4673-7121-6/15/$31.00 ©2015 IEEE

[17] Osama Abdel-Raouf , Mohamed Abdel-Baset and Ibrahim El-henawy ,” An Improved Flower Pollination Algorithm with Chaos”, in: I.J. Education and Management Engineering, 2014, 2, 1-8.

[18] Osama Abdel-Raouf, Ibrahim El-henawy, Mohamed Abdel-Baset," A Novel Hybrid Flower Pollination Algorithm with Chaotic Harmony Search for Solving Sudoku Puzzles", in: I.J. Modern Education and Computer Science, 2014, 3, 38-44.

[19] Mohamed Abdel-Baset, Ibrahim M. Hezam," An Effective Hybrid Flower Pollination and Genetic Algorithm for Constrained Optimization Problems" , in: Adv. Eng. Tec. Appl. 4, No. 2, 27-34 (2015).

[20] O. Abdel-Raouf, M. Abdel-Baset, I. El-henawy," A New Hybrid Flower Pollination Algorithm for Solving Constrained Global Optimization Problems", in: International Journal of Applied Operational Research Vol. 4, No. 2, pp. 1-13, Spring 2014.

[21] Dwaipayan Chakraborty, Sankhadip Saha, Oindrilla Dutta," DE-FP A: A Hybrid Differential Evolution-FlowerPollination Algorithm for Function Minimization", 978-1-4799-5958-7114/$31.00 102014 IEEE.

[22] R.Jensi, G.Wiselin Jiji," HYBRID DATA CLUSTERING APPROACH USING K-MEANS AND FLOWER POLLINATION ALGORITHM", in: Advanced Computational Intelligence: An International Journal (ACII), Vol.2, No.2, April 2015.

[23] Hari Mohan Dubey, ManjareePandit, B.K. Panigrahi," Hybrid flower pollination algorithm with time-varying fuzzy selection mechanism for wind integrated multi-objective dynamic economicdispatch", 0960-1481/© 2015 Elsevier Ltd.

[24] NavdeepKaurJohal, Samandeep Singh, Harish Kundra," A hybrid FPAB/BBO Algorithm for Satellite ImageClassification", in: International Journal of Computer Applications (0975 – 8887)Volume 6– No.5, September 2010.

[25] Emad Nabil," A Modified Flower Pollination Algorithm for Global Optimization", Article in Expert Systems with Applications · April 2016.

[26] Mohamed Abdel-Baset, Ibrahim Hezam," A Hybrid Flower Pollination Algorithm for Engineering Optimization Problems", in: international Journal of Computer Applications (0975 – 8887) Volume 140 – No.12, April 2016.

[27] Ibrahim M. Hezam, Mohamed Abdel-Baset, Bayoumi M. Hassan," A Hybrid Flower Pollination Algorithm with Tabu Search for Unconstrained Optimization Problems", Inf. Sci. Lett. 5, No. 1, 29-34 (2016).

[28] GaganpreetKaur, Dheerendra Singh, ManjinderKaur," Robust and Efficient 'RGB' based Fractal Image Compression: Flower Pollination based Optimization", in: International Journal of Computer Applications (0975 – 8887) Volume 78 – No.10, September 2013.

[29] MarwaSharawi, E. Emary, ImaneAlySaroit, Hesham El-Mahdy," Flower Pollination Optimization Algorithm forWireless Sensor Network Lifetime GlobalOptimization", in: International Journal of Soft Computing and Engineering (IJSCE)ISSN: 2231-2307, Volume-4, Issue-3, July 2014.

[30] E. Emary, Hossam M. Zawbaa, Aboul Ella Hassanien, Mohamed F. Tolba and Vaclav Snasel," Retinal vessel segmentation based on flower pollination search algorithm".

[31] MeriemBensouyad," A discrete flower pollination algorithm for graph coloring problem", Cybernetics (CYBCONF), 2015 IEEE 2nd International Conference on, Page(s):151 – 155,Print ISBN: 978-1-4799-8320-9.

[32] Wang, Rui; Zhou, Yongquan; Zhou, Yuxiang; Bao, Zongfan," Local Greedy Flower Pollination Algorithm for Solving Planar Graph Coloring Problem", in: Journal of Computational and Theoretical Nanoscience, Volume 12, Number 11, November 2015, pp. 4087-4096(10).

[33] Khalil AL-Wagih," An Improved Flower Pollination Algorithm for Solving Integer Programming Problems", in: Appl. Math. Inf. Sci. Lett. 3, No. 1, 31-37 (2015).

[34] Ibrahim El-henawy, Mahmoud Ismail," An Improved Chaotic Flower Pollination Algorithm for Solving LargeInteger Programming Problems",in: International Journal of Digital Content Technology and its Applications(JDCTA)Volume8, Number3, June 2014.

[35] Mohamed Abdel-Baset, Ibrahim M. Hezam," An Improved Flower Pollination Algorithm for Ratios Optimization Problems", in: Appl. Math. Inf. Sci. Lett. 3, No. 2, 83-91 (2015).

[36] RuiWang and Yongquan Zhou," Flower Pollination Algorithm with Dimension by Dimension Improvement", in: Hindawi Publishing Corporation Mathematical Problems in EngineeringVolume 2014, Article ID 481791, 9 pages.

[37] yongquan Zhou, Rui Wang," An Improved Flower Pollination Algorithm for Optimal UnmannedUndersea Vehicle Path Planning Problem", Article in International Journal of Pattern Recognition and Artificial Intelligence · January 2016.

[38] Alberto Ochoa, Saúl González, Lourdes Margain, Teresa Padilla, Oscar Castillo, Patricia Melín," Implementing Flower Multi-objective Algorithm for selection of University academiccredits", August 2015.

[39] S. Sakthivel, P. Manopriya, S. Venus, S. Ranjitha and R. Subhashini," Optimal Reactive Power Dispatch Problem Solved By Using Flower Pollination Algorithm", in: International Journal of Applied Engineering Research ISSN 0973-4562 Volume 11, Number 6 (2016) pp 4387-4391.

[40] A. Sudheer K umarDr.A.V. `Giridhar," A New Meta Heuristic Algorithm Based ShuntCapacitive Compensation for Power LossReduction on Radial Distribution System", in: International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181Vol. 3 Issue 10, October- 2014.

[41] C. Shilaja, K. Ravi," Optimal Line Flow in Conventional Power System using Euclidean Affine Flower Pollination Algorithm", in: international journal of renewable energy research C.Shilaja and K.Ravi, Vol.6, No.1, 2016.

[42] PrernaSaxena and Ashwin Kothari," Linear antenna array optimization using flower pollination algorithm", Saxena and Kothari SpringerPlus (2016) 5:306.

[43] Mohammad Musofa Mulya Pambudy, Sasongko Pramono Hadi, Husni Rois Ali," Flower Pollination Algorithm for Optimal Control inMulti-Machine System with GUPFC", 2014 6th in: International Conference on Information Technology and Electrical Engineering (ICITEE), Yogyakarta, Indonesia.

[44] V. S. S. S. Chakravarthy Vedula,S. R. Chowdary Paladuga, and M. Rao Prithvi," Synthesis of Circular Array Antenna for Sidelobe Level and Aperture Size Control Using Flower Pollination Algorithm", in: Hindawi Publishing Corporation International Journal of Antennas and Propagation Volume 2015, Article ID 819712, 9 pages.

[45] Priyanka Das and Manoj Kumar ," Improvement of Voltage Profile using Shunt Capacitor in Radial Distribution System by Flower Pollination Algorithm and Sensitivity Analysis", in: Journal of Basic and Applied Engineering Research p-ISSN: 2350-0077; e-ISSN: 2350-0255; Volume 2, Number 18; July-September, 2015, pp. 1555-1559.

[46] Priyanka Das and Manoj Kumar," Voltage Profile Improvement using Distributed Generation in 33-bus and 69-bus RDS by Flower Pollination Algorithm and Sensitivity Factor", in: Advanced Research in Electrical and Electronic Engineering p-ISSN: 2349-5804; e-ISSN: 2349-5812 Volume 2, Issue 14 October-December, 2015, pp. 37-41.

[47] A.Y. Abdelaziz , E.S. Ali, S.M. AbdElazim ," Optimal sizing and locations of capacitors in radial distribution systems via flower pollination optimization algorithmand power loss index", in: Engineering Science and Technology, an International Journal 19 (2016) 610–618.
[48] Dinakara Prasad Reddya,, V.C. Veera Reddy, T. Gowri Manohara," Application of flower pollination algorithm for optimal placement and sizing of distributed generation in Distribution systems", in: Journal of Electrical Systems and Information Technology ARTICLE IN PRESS
[49] Eyad S. Oda , Abdelazeem A. Abdelsalam , Mohamed N. Abdel-Wahab ,Magdi M. El-Saadawi ," Distributed generations planning using flowerpollination algorithm for enhancing distributionsystem voltage stability", in: Ain Shams Engineering Journal (2015) ARTICLE IN PRESS .
[50] D. Lakshmi, A. Peer Fathima, Ranganath Muthu," A Novel Flower Pollination Algorithm to Solve Load Frequency Control for aHydro-Thermal Deregulated Power System", Circuits and Systems, 2016, 7, 166-178 Published Online April 2016 in SciRes.
[51] A.Y. Abdelaziz , E.S. Ali , S.M. AbdElazim ," Flower pollination algorithm to solve combined economic and emission dispatch problems", in: Engineering Science and Technology, an International Journal ARTICLE IN PRESS .
[52] R. Prathiba, M. Balasingh Moses, S. Sakthivel," Flower Pollination Algorithm Applied forDifferent Economic Load Dispatch Problems", in: International Journal of Engineering and Technology (IJET) ISSN : 0975-4024 Vol 6 No 2 Apr-May 2014.
[53] Hari Mohan Dubey, B.K. Panigrahi, and ManjareePandit," Improved Flower Pollination Algorithm for Short Term Hydrothermal Scheduling", © Springer International Publishing Switzerland 2015B.K. Panigrahi et al. (Eds.): SEMCCO 2014, LNCS 8947, pp. 721–737, 2015.
[54] Udjianna S. Pasaribu, Fathimah al-Ma'shumah, DonyPermana," Estimation of the Transition Matrix in Markov Chain Model of Customer Lifetime Value Using Flower Pollination Algorithm", in: Applied Mathematical Sciences, Vol. 9, 2015, no. 69, 3409 – 3418.
[55]Sinan Melih Nigdeli, Gebrail Bekdaş, Xin-She Yang," Application of the Flower Pollination Algorithm in Structural Engineering", Volume 7 of the series Modeling and Optimization in Science and Technologies pp 25-42.

International Journal of Computer Applications Technology and Research
Volume 5– Issue 6, 347 - 352, 2016, ISSN:- 2319–8656

Design of High Speed Phase Frequency Detector in 0.18 µm CMOS Process for PLL Application

Prof. Abhinav V. Deshpande

Assistant Professor

Department of Electronics & Telecommunication Engineering

Prof. Ram Meghe Institute of Technology & Research, Badnera, Amravai-444701, Maharashtra India

Abstract:

The Phase Frequency Detectors (PFD's) are proposed in this research paper by using the two different structures of D Flip-Flop that is the traditional D Flip-Flop and modified D Flip-Flop with a NAND gate which can overcome the speed and area limitations of the conventional PFD. Both of the PFD's use 20 transistors. The traditional PFD consumes 133.92 µW power when operating at 40 MHz frequency with 1.8 Volts supply voltage whereas the modified PFD consumes 100.51 µW power operating at 40 MHz frequency with 1.8 Volts supply voltage. The designs are implemented by using 0.18 meter CMOS process in Tanner 13.ov. These can be used in PLL for high speed applications.

Keywords: CMOS, Phase Locked Loop (PLL), D Flip-Flop, Phase Frequency Detector (PFD), NAND gate, Clock Signal

1. INTRODUCTION:

PFD is one of the main blocks of the PL which is used in various applications such as Wireless Communication systems. Digital Circuits, Sensor Receivers etc. The implementation of a fully integrated, low power and high performance PLL significantly affect the overall system performance [1][2].

In the PLL, the Phase Frequency Detectors (PFD's) compares the rising edges of the reference clock and the voltage controlled oscillator (VCO) clock, and generates a lead signal when the reference phase is leading or a lag signal when the reference phase is lagging [2][3]. The phase difference which is detected in the PFD passes through the loop filter to control the VCO.As the phase difference critically affects the overall characteristics of the PLL such as the lock in time and jitter performance, the PFD should be designed in order to work accurately for any phase difference [1][2][4].

The design of PFD consists of two flip-flops and NAND gate to provide a reset path [3][4]. As shown in Figure 1, the D input of the flip-flops is connected to high and the input signals are applied to the clock input. When one of the clock changes to high, this flip-flop will be charged and change its output to high [4][5]. The NAND gate is for preventing both the flip-flops to be high at the same time. As we can see the inputs of the NAND gate are both Up and Down signal from both the flip-flops and the output of the NAND gate is connected to the

reset input of the flip-flops. As soon as both the outputs (Up and Down) are high the NAND gate will generate a high signal that will reset both flip-flops by avoiding the situation of both high at the same time [5][7].

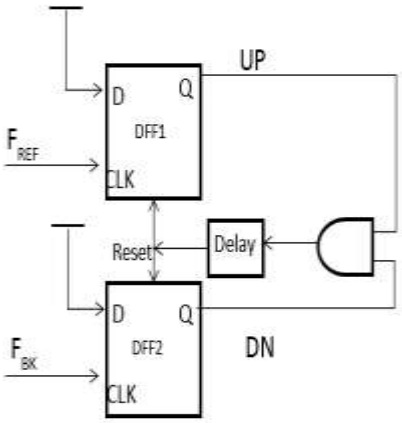


Figure 1 Block Diagram of PFD

The following figures Figure 2 and Figure 3 shows the schematic diagram of traditional and the modified D Flip-Flop respectively which is used to design the architecture of traditional and modified PFD [6][7].

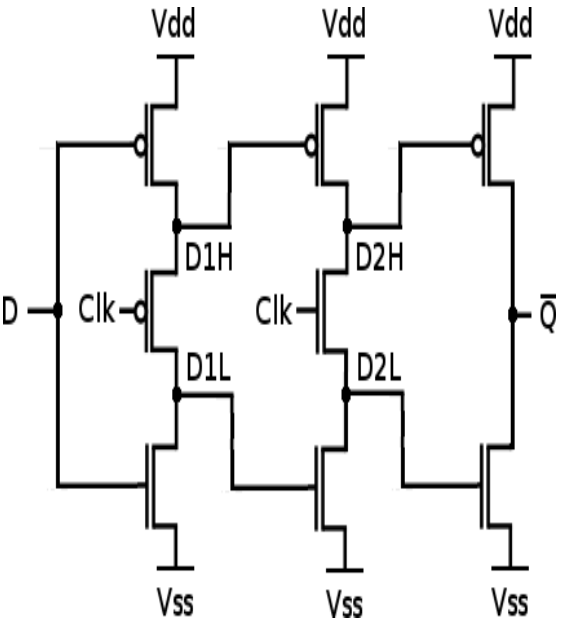


Figure 2 Schematic Diagram of Traditional D Flip-Flop

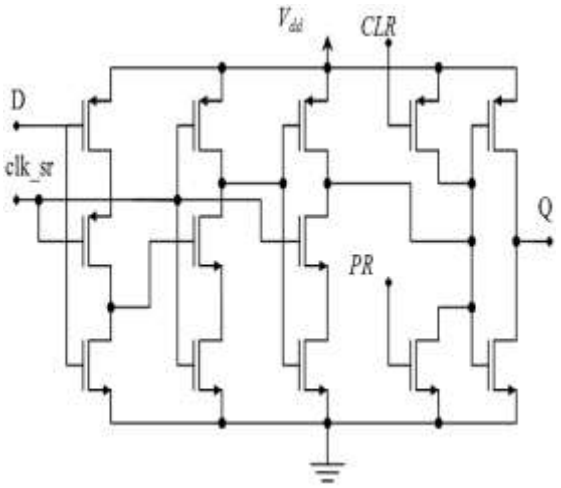


Figure 3 Schematic Diagram of Modified D Flip-Flop

2. TRADITIONAL PHASE FREQUENCY DETECTOR:

This research paper presents two PFD architectures having low area and can work on higher frequencies [7][8]. Figure 4 shows the Phase Frequency Detectors (PFD's by using NAND gate). The circuit consists of two resettable, edge triggered traditional D Flip-Flops with their D inputs tied to logic 1 [6][7]. The CLK1 and CLK serve as the clocks of the flip-flops. Suppose the rising edge of CLK1 leads that of CLK, then UPb goes to logic high. UPb keeps high until the rising edge of the CLK makes DNb on high level. Because UPb and DNb are NANDed, so RESET goes to logic high and resets the PFD into the initial state [6][7][8]. The schematic of NAND gate based PFD circuit consisting of only 20 transistors is as given in Figure 4.

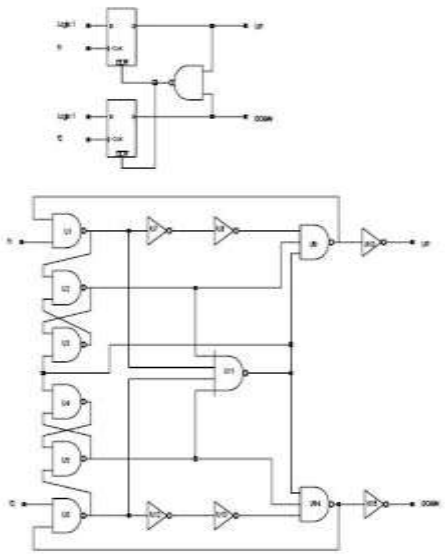


Figure 4 Schematic Diagram of Traditional PFD by using NAND Gates

3. MODIFIED PHASE FREQUENCY DETECTOR:

Figure 5 shows the phase frequency detector by using NAND gate. The circuit consists of two resettable, edges triggered D flip-flops with their D inputs tied to logic 1. The CLKREF and CLK serve as the clocks of the flip-flops. The UPb and DNb signals are given as input to the NAND gate. Suppose the rising edge of CLKREF leads that of CLK, then UPb goes to logic low i.e. Up keeps high until the rising edge of CLK makes DNb on low level [7][8]. Because UPb and DNb are NORed, so RESET goes to logic high and resets the PFD into the initial state. The circuit is implemented by using 0.18 µm CMOS process in Tanner 13.0v with only 20 transistors.

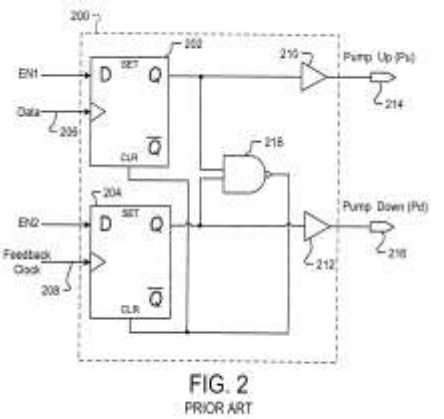


FIG. 2
PRIOR ART

Figure 5 Schematic Diagram of Modified PFD by using NAND Gates

4. SIMULATION AND RESULTS:

Both the PFD circuits are simulated on Tanner 13.0 at 1.8Volts in order to obtain the results

with the input frequency of 40 MHz [2][3][8]. The NAND gate based PFD circuit is simulated on Tanner 13.0 at 1.8 Volts in order to obtain the results with the input frequency of 40 MHz as shown in Figure 6.

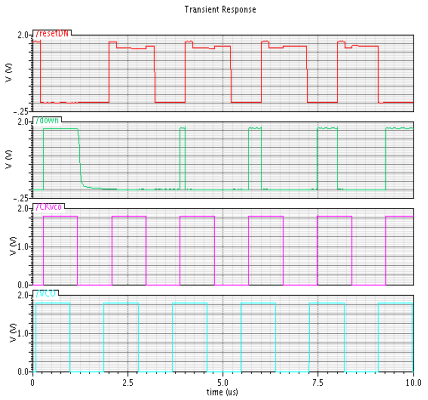


Figure 6 Waveforms of NAND Gate-Based Traditional PFD

The NAND gate based modified PFD circuit is simulated on Tanner 13.0 at 1.8 Volts in order to obtain the results with the input frequency of 40 MHz as shown in Figure 7.

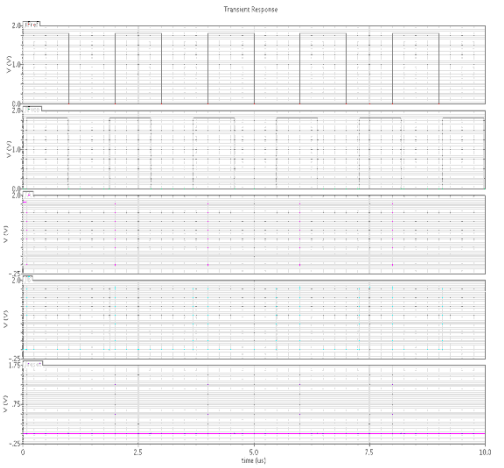


Figure 7 Waveforms of NAND Gate-Based Modified PFD

The two circuits can be simulated when the Free and Face have different frequencies [3][4][5]. Figure shows the waveform for NAND gate based PFD. The Up signal and the DN signal go high at the rising edge of Free and Face. When both Up and Dn signals become logic high the circuit is reset to the initial state. The pulse width of the UP and DN signal is proportional to the frequency difference between the two inputs [4][5][7].

5. PERFORMANCE AND COMPARISON:

Type of PFD	Operating Frequency	No. of Transistors	Power Consumption	Delay	Dead Zone
Traditional PFD	40 MHz	20	133.91 µW	10 ns	9 ps
Modified PFD	40 MHz	20	100.51 µW	10 ns	1 ps

6. CONCLUSION:

This research paper presents two PFD designs which are implemented in the 0.18 µm CMOS process. Both of the PFD consists of only 20 transistors and can operate up to 1 GHz frequency but the modified PFD preserves the main functionality of traditional PFD with low power consumption.. The dead zone of traditional PFD is 9 ps whereas for modified PFD is 6 ps. The performance of the two PFD's is compared against the traditional PFD in Table 1.

7. ACKNOWLEDGMENTS:

I am very much thankful to all of the staff members and the Head of Department, Electronics & Telecommunication Engineering, Prof. Ram Meghe Institute of Technology & Research, Badnera, Amravati-444701 for their kind support and co-operation in successful carrying out this research work. This research work was undertaken as a part of Technical Education Quality Improvement Program (TEQIP-2) in order to promote and facilitate the current and emerging trends in the field of Electronics & Telecommunication Engineering so that the new and young researchers working in the fields of research and development in Electronics Engineering domain should get the benefit of pursuing their main hobbies which are pertaining to the Embedded Systems platform and should try to learn the new skills and expertise in the particular field of Embedded Systems and Wireless Networks.

8. REFERENCES:

[1] Mohammad Zaher Al Sabbagh B. S., "0.18 µm Phase/Frequency Detector and Charge Pump Design for Digital Video Broadcasting for Handheld Phase Locked Loop Systems", THESIS, The Ohio State University, 2007.

[2] W. Hu et. al., "Fast Frequency Acquisition Phase Frequency Detector with Zero Blind Zones in PLL", In IEEE Electronics Letters, Volume 43, No. 19, September 2007.

[3] H. O. Johanson, "A Simple Precharged CMOS Phase Frequency Detector", In IEEE Journal of Solid State Circuits, Volume 33, No. 2, February 1998, pp. 295-299.

[4] W. H. Lee, J. D. Cho and S. D. Lee, "A High Speed and Low Power Phase-Frequency Detector and Charge-Pump", In Proceedings of Asia South Pacific Des. Autom., Conference, January 1999, Volume 1, pp. 269-272.

[5] W. H. Chen, M. E. Inerowicz and B. Jung, "Phase Frequency Detector with Minimal Blind Zone for Fast Frequency Acquisition", IEEE Transactions on Circuits and Systems-II:Express Briefs, Volume 57, No. 12, December 2010.

[6] N. M. Hammam Ismail and M. Othman, "CMOS Phase Frequency Detector for High Speed Applications", 978-1-4244-5750-2/10/2009 IEEE.

[7] B. Razavi, "Design of Analog CMOS Integrated Circuits", Prentice Hall, 1998.

[8] R. Jacob Baker, "CMOS Circuit Design, Layout and Simulation".

International Journal of Computer Applications Technology and Research
Volume 5– Issue 6, 353 - 357, 2016, ISSN:- 2319–8656

Congestion Control in Wireless Sensor Networks Using Genetic Algorithm

Navid Samadi Khangah
Department of Computer Engineering, Tabriz Branch, Islamic Azad University Tabriz, Iran

Ali Ghaffari
Department of Computer Engineering, Tabriz Branch, Islamic Azad University Tabriz, Iran

Abstract: Sensor network consists of a large number of small nods, strongly interacting with the physical environment, takes environmental data through sensors, and reacts after processing on information. Wireless network technologies are widely used in most applications. As wireless sensor networks have many activities in the field of information transmission, network congestion cannot be thus avoided. So it seems necessary that some new methods can control congestion and use existing resources for providing better traffic demands. Congestion increases packet loss and retransmission of removed packets and also wastes of energy. In this paper, a novel method is presented for congestion control in wireless sensor networks using genetic algorithm. The results of simulation show that the proposed method, in comparison with the algorithm LEACH, can significantly improve congestion control at high speeds.

Keywords: wireless sensor networks, congestion control, genetic algorithm, optimization, clustering.

1. INTRODUCTION

Wireless sensor networks are networks including independent sensors in environment which measure physical or environmental conditions such as temperature, sound, vibration, pressure, motion or pollutants in different locations. These sensors are small and work interacting with another and have a limited amount of stored energy, amount of memory and bandwidth. Restrictions such as buffer memory, limited computing, stored electric power have caused to be proposed several methods for routing and data transmission. Congestion in wireless sensor networks occurs when sensor nodes receive more number of packets than the number they can send. Therefore, it is necessary to use fast and efficient congestion control mechanisms in wireless sensor networks [1]. Therefore, transport layer in wireless sensor networks is responsible for controlling congestion. Congestion control methods are done by the two methods: traffic and resource control. Congestion not only causes loosing severe information, but also leads to excessive consumption of energy in the nodes. Many multicast routing protocols are provides for data transmission in multiple paths, however congestion control mechanisms are rarely to be found for multiple routing. In this paper, a congestion control protocol is provided for wireless sensor networks using genetic algorithms aimed at increasing reliability and lifetime level of the network and the high throughput.

2. CONGESTION CONTROL

There are generally two reasons for Congestion in wireless sensor networks. The first reason is that packet arrival rate is much more of packet service rate. This mode happens more to nodes near the destination node, because they usually carry more upward combined traffic. The second reason is due to aspects of performance in level link such as competition, interference and bit error rate. Congestion in wireless sensor networks has a direct impact on energy efficiency and quality of service. For example, congestion can lead to a buffer overflow, hence to a larger queue delay and losing more packets. Packet loss not only leads to reduce reliability and quality of service but also squandering the nodes limited energy. Thus, congestion in sensor networks should be controlled efficiency by avoiding congestion occurring, or reducing the congestion. Several congestion control techniques are proposed for wireless sensor networks [2][3][4]. All the congestion control mechanisms have the same basic goal: All of them first try to diagnose or detect congestion, and after detection of congestion are too aware other groups of the situation of congestion. In general, there are three phases for congestion control: Congestion detection phase, Congestion notification phase and rate adjustment phase [5][6].

3. RELATED WORKS

Generally different protocols are introduced for transport layer in wireless sensor networks so that each is able to effectively control Congestion. Among the proposed protocols, there are some protocols that control just congestion controlling and some guarantee just reliability. There are rare number of these protocols which can both ensure and control the reliability and congestion.

SRCP (Sensor Reliability and Congestion Control Protocol) algorithm determines traffic, based on increasing the packet send process and decreases throughput in a short time. This is a rate based protocol which adjusts distance of sending packets after a fixed time, called track time [7]. Mesh interface algorithm, as network topology input, uses desired sent rate flow for routing directions. Imaging the network interference, as a conflict diagram, is approximate and dependent using an iterative process in order to estimate (fair max-min) secure transmission rate for each flow to reflect the total network throughput [8]. SPEED algorithm tries to maintain the desired speed in real time and does the tasks uniformly by providing applications. Traffic diversion causes end-to-end delay to be proportional to the distance between the source and destination through multiple paths and adjust the transmission rate [9]. HTAP (Hierarchical Tree Alternative Path) algorithm tries to guarantee reliability of applications during the period of overloading (overload) without reducing the funds rate at the time of important events. HTAP which is a combination of two algorithms: creation of alternative route, creation of a hierarchical tree, chooses it by using the network congestion [10]. TARA (Thermal-Aware Routing Algorithm) algorithm defines total link congestion to be as the total of traffic and traffic interference links and selects bottleneck the large amounts of congestion as well [11]. WCCP (WMSN Congestion Control Protocol) algorithm is a two-part protocol. In its resources sector, SCAP is used to set transmission rate and distribution of abandoned pockets. The aim of SCAP

protocol, at the first, is to avoid the congestion from the source node. In receiver, RCCP protocol is used for detecting the congestion occurrence and informing the source nodes in the regarding congestion [12]. LEACH (Low Energy Adaptive Clustering Hierarchy protocol) is self-organizing clustering protocol which distributes energy load on the network sensors. In LEACH, nodes organize themselves into local clusters so that a node acts in the cluster as the cluster head [13].

4. THE PROPOSED ALGORITHM

Before a genetic algorithm can be executed first must be found a suitable encoding (or representation) for intended problem intended. The usual manner of representing chromosomes in genetic algorithm is a binary string. Each decision variable becomes a binary form and then is created chromosome by getting together the variables. Although this method is a method of encoding extended but other ways are growing such as representation by real numbers. A fitness function must be also devised to attribute a value to each coded solution. During the run, parents are selected for breeding and, using crossover or mutation operators, are combined together to produce new children. This process is repeated several times to produce the next generation population. Then this population is investigated and if the convergence criteria are met, the above process is terminated. Work process in genetic algorithm is done by two mutation and integration methods. Work process of integration method is studied in the proposed algorithm as follows.

First the nodes coordinate are named and the number of nodes is identified. The number of nodes is defined as nnodes. Then the nodes are put in a variable named$k$. The variable k is divided into two parts k_1 andk_2. By selecting the number of nodes, an amount of primary energy is given to each of them (initial amount of energy is between 0 and 1). After this step in every time of overall execution, genetic algorithm is executed as much as 10 times (the number 10 is the numerical value of the option).

4.1 Stages of integration implementation

Coordinate nodes are considered 8-bit, k_1 and k_2 are the integer numbers converted to binary and placed in new variables k_3 and k_4. k_3 and k_4 are the same chromosomes in the proposed algorithm. At this stage the variables C, D are defined in order to locate chromosomes which ultimately are obtained from creation of chromosomes positions (coordinate) of nodes. To perform the integration, the cut takes place randomly into chromosomes in each time intercourse so that the state spatial coordinates (chromosomes) are improved by this process. have a 9-point text, as you see here. Please use sans-serif or non-proportional fonts only for special purposes, such as distinguishing source code text. If Times Roman is not available, try the font named Computer Modern Roman. On a Macintosh, use the font named Times. Right margins should be justified, not ragged.

4.2 Stages of implementation mutation

There is a variable called m in mutation randomly chosen to practice jumps and again the chromosomes are placed in variables A and$B$. By examining$m$, if $Am = 0$, the values one are put at chromosome home, otherwise zero. Mutation output will be as integration in C and D. After combination and mutation, cluster operations are carried out. For selecting cluster head, the amount battery is used, while for proper functioning, the position of the nodes, i.e., the node distance. The formular for selecting cluster head is as follows:

$$\text{Temp} = \frac{p}{1 - p^* \left(r \bmod \frac{1}{p}\right)}$$

Where p is the probability cluster head node is and placed equal 0.1 and r the current stage.

In case of holding the cluster head, for calculating the distance of the nodes coordinates from the sink, equation (2) is used [14]:

$$\text{distance} = \sqrt{(E_i - E_{nnodes+1})^2 + (F_i - F_{nnodes+1})^2}$$

where E_i, E_{nnodes}, F_i, F_{nnodes} denote, respectively, coordinates of nodes, number of nodes in x axis, the coordinates of nodes, and the number of nodes in y axis. Before the genetic operations, a random amount of primary energy is assigned to each node. Here (in GA) the distance of the node to sink is initially examined for required processing (operation) in order to assign the amount of energy which a node needs in a specified distance between the node and sink. After determining the amount of primary energy for each node in the new population, the aim of the proper functioning of the proposed algorithm is described in the sequel. Nodes distance ratio to sink and primary energy nodes (obf) is obtained by equation (3):

$$\text{obf} = \sqrt[2]{\frac{(E_i - E_{nnodes+1})^2 + (F_i - F_{nnodes+1})^2}{\text{Intionl}E(1,i)}}$$

where E_i, E_{nnodes}, F_i, F_{nnodes}, are as before, $IntionlE(1,i)$ is initializing in matrix $1 \times i$ and i is the number of nodes. All of the values obtained by the above ratios are placed in a variable called sum. To the number of nodes, a new variable, named prob, in order to accommodate the matrix of relationships which can save matrix values is established in prob variable. Appropriate matrix values are good in sum of the values. The first entry of the prob matrix is stored in prob variable, as a minimum, to be designed by finding prob matrix global minimum so that in the end, according to the above studies, a new state space (coordinates) is established for sink.

5. THE IMPLEMENTATION RESULTS

In this section the results obtained by the implementation of the proposed method are evaluated. MATLAB software is used for evaluating the proposed method whereas LEACH algorithm for comparing. Simulation parameters are specified in Table 1.

Table 1. Table captions should be placed above the table

Parameters	Value
Number of nodes	50, 100, 150, 200
Environment size	100m*100m
Sink localization	(50*50)
Energy model	0.5 J
Rounds	500, 1000, 1500, 2000

The state space of nodes is displayed as in Figures 2-5 in LEACH and LECH_GA algorithms, respectively.

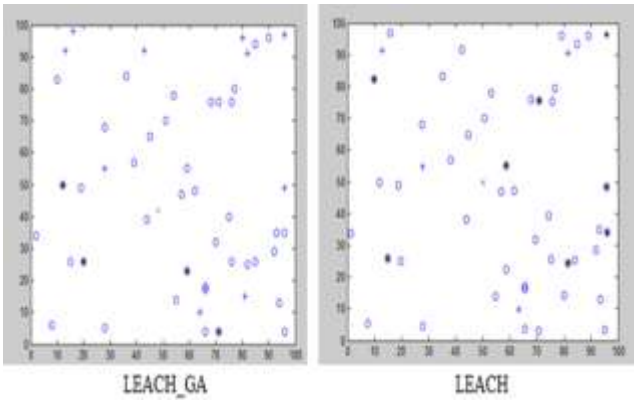


Figure. 2 The state space for LEACH and LECH_GA algorithms with the number of 50 nodes and 500 rounds

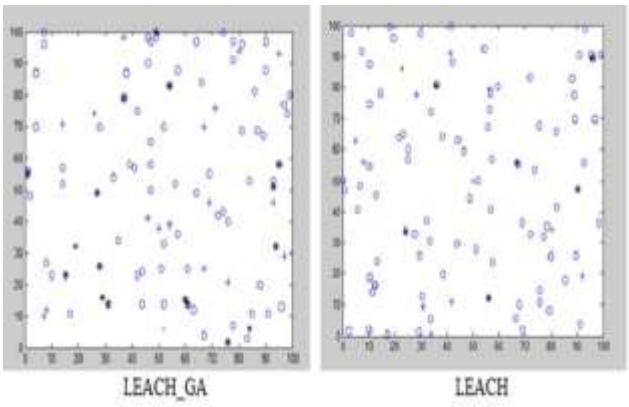


Figure. 3 The state space for LEACH and LECH_GA algorithms with the number of 100 nodes and 1000 rounds

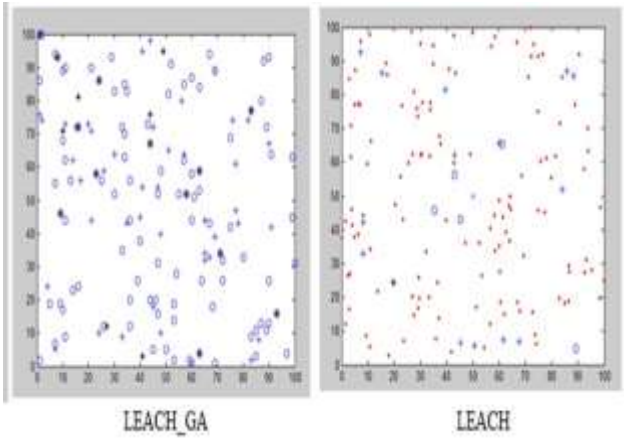


Figure. 4 The state space for LEACH and LECH_GA algorithms with the number of 150 nodes and 1500 rounds

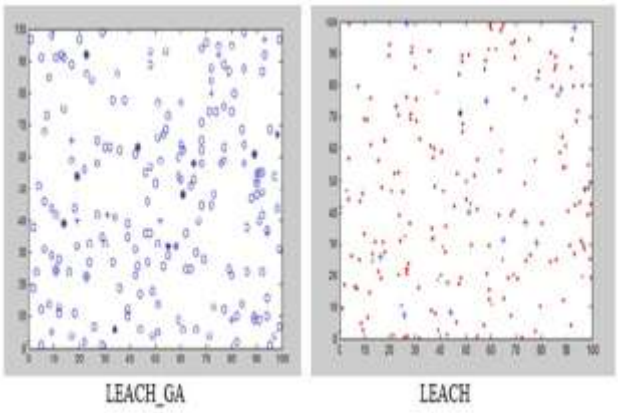


Figure. 5 The state space for LEACH and LECH_GA algorithms with the number of 200 nodes and 2000 rounds

Figure 6 compares the dead nodes for (LEACH GA) and (LEACH) algorithms. In around 500, the number of dead nodes in both algorithms is zero. In around 1000 onwards, the dead nodes in LEACH algorithm dramatically increased, but this amount is zero for LECH_GA algorithm. Due to the high number of rotations, the implementation time of the algorithm becomes longer and dead nodes largely increase in LEACH algorithm, and genetic algorithms due to having more energy, still do not have any dead nodes.

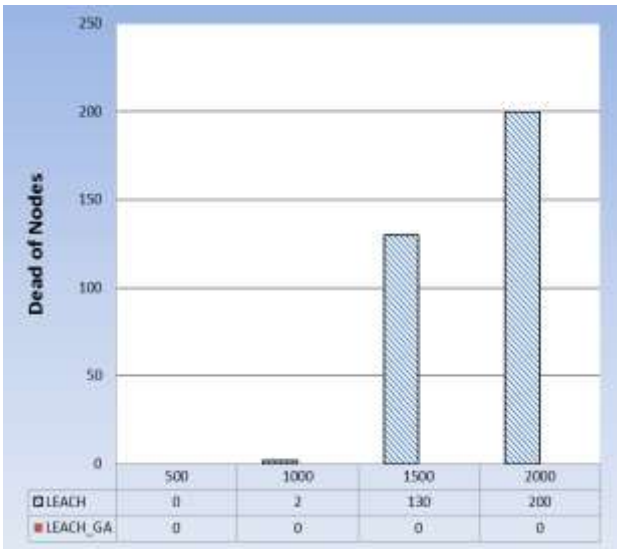


Figure. 6 Comparison of dead nodes in the proposed algorithm and LEACH algorithm

Figure 7 shows the comparison of ripe nodes to destination (base station) in the two genetic algorithms LECH_GA and LEACH. According to Figure 7, in period 500 the number of incoming packets in LEACH and LECH_GA algorithms are respectively 2508 and 7412. By increasing the number of rounds, the number of packets received to the destination will also increase. This is due to changes in sink state space, and the position of the nodes by the proposed method and there is a significant difference in the number of rotations. On the other hand in LEACH algorithm by increasing the number of turns, live nodes decrease and markedly decrease the packets ripen to the destination.

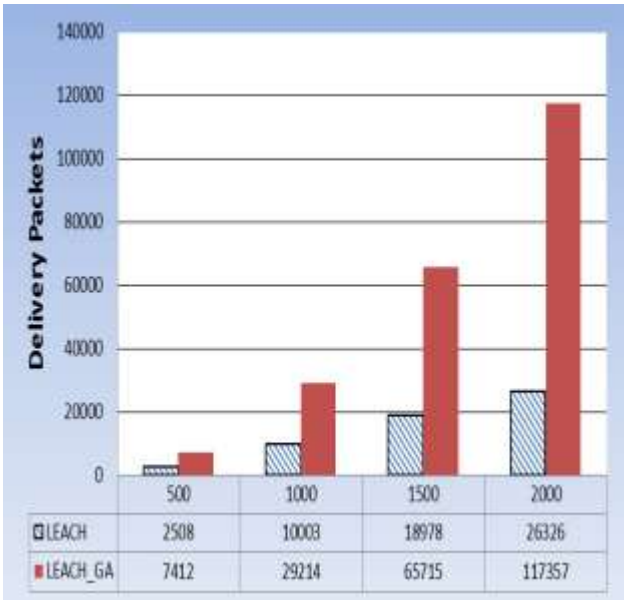


Figure. 7 The number of incoming packets to sink

Figure 8 shows the comparison of the number of choosing cluster in the two genetic algorithms LECH_GA and LEACH. In lower revs in the genetic algorithm, no need to select much cluster, according to the production chromosomes. Also it is observed that by increasing the rounds, due to high energy consumption, cluster head with high reliability is chosen for packet transmission in genetic algorithms.

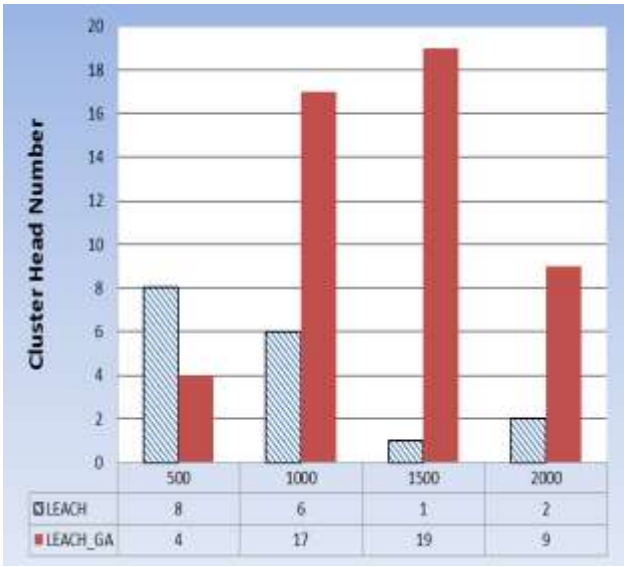


Figure. 8 Selection of cluster head

Figure 9 shows the comparison of energy consumption for genetic algorithm, LEACH-GA and LEACH. Energy consumption for the proposed algorithm has not so declined, but energy consumption for LEACH algorithm has ended in a round 2000.

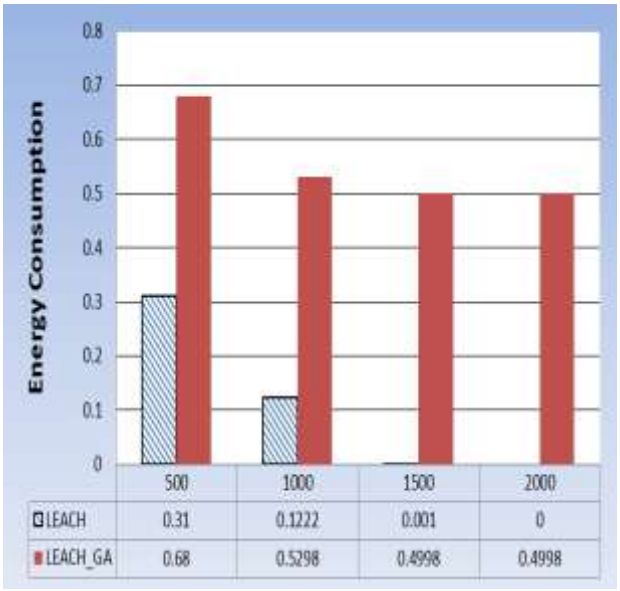


Figure. 9 The remaining energy consumption in both LEACH and LEACH_GA algorithms

6. CONCLUSION

In this article, congestion control was studied using genetic algorithms in wireless sensor networks. According to the results, it was given a new value for the state space of the sink so that it can have good performance at sending incoming packets to the destination (sink) when the position (state-space) of the sink changes, reducing nodes losses, traducing energy consumption power, selecting better cluster. Whatever the number of the terms is more in the proposed algorithm, the performance in terms of dead nodes, number of incoming packets to the destination and energy would be highly better. A deep understanding of congestion control lets the transmission comprehensive protocol design with reliability. Addressing the transport layer with high reliability is essential to ensure effectiveness of the applications. Checking the transfer protocols shows controls of the reliability of a vision to this task (congestion control).

7. REFERENCES

[1] Akyildiz, I. F., Su, W., Sankarasubramaniam, Y. 2002. Computer Networks: The International Journal of Computer and Telecommunications Networking archive Volume 38 Issue 4, 393-422.

[2] Tao, Q. and Yu, F. Q. 210. ECODA: Enhanced Congestion Detection and Avoidance for Multiple Class of Traffic in Sensor Networks. in proc. IEEE conf.

[3] Vijayaraja, V. and Hemamalini, R. Congestion in Wireless Sensor Networks and Various Techniques for Mitigating Congestion - A Review. in proc. IEEE conf. May2010.

[4] Sridevi, S. Usha, M. and Pauline, G. 2012. Priority Based CongestionControl For Heterogeneous Traffic In Multipath Wireless Sensor Networks. inproc. IEEE conf.

[5] Hull, b. jamieson, k. and Balakrishnan, h. 2004. *Mitigation congestion in wireless sensor networks.* in procedding of ACM sensys, vol.4, USA,

[6] Wang, C. and Sohraby, K. 2006. *A Survey of Transport Protocols for Wireless Sensor Networks.* IEEE Network, vol.20, pp.34-40,

[7] Tezcan, N., and Wang, W. 2007. ART: an asymmetric and reliable transport mechanism for wireless sensor networks. Int. J. Sen. Netw, 2(3), 188–200.
[8] Sundaresan, K., Anantharaman, V., Hsieh, H. Y., Sivakumar, R. 2005. ATP: A Reliable Transport Protocol for Ad Hoc Networks. IEEE Trans. Mobile Computing, 4(6), 588–603.
[9] Kang, J., Zhang, Y., Nath, B. 2007. TARA: Topology-Aware Resource Adaptation to Alleviate Congestion in Sensor Networks. IEEE Trans. Parallel Distrib. Syst, 18(7), 919–931.
[10] Wang, G., & Liu, K. 2009. Upstream hop-by-hop congestion control in wireless sensor networks. in Proc. IEEE 20th Int. Symp. on Personal, Indoor and Mobile Radio Communications, 1406–1410.
[11] Monowar, M., Rahman, M., Hong, V. S. 2008. Multipath Congestion Control for Heterogeneous Traffic in Wireless Sensor Network, in Proc. 10th Int. Conf. on Advanced Communication Technology, ICACT, 3, 1711–1715.
[12] Mahdizadeh Aghdam, S., Khansari, M. , Rabiee, H. R., Salehi, M. 2014. WCCP: A congestion control protocol for wireless multimedia communication in sensor networks. Ad Hoc Networks 13 , 516–534.
[13] Gou, H., Yoo, Y., Zeng, H. 2009. A Partition-Based LEACH Algorithm for Wireless Sensor Networks. CIT '09. Ninth IEEE International Conference on, Computer and Information Technology, 2009, IEEE, PP. 40 – 45.
[14] Feigenbaum, J., Kannan, S., McGregor, A., Suri, S. and Zhang, J. 2005. Graph distances in the streaming model: the value of space. In ACM-SIAM Symposium on Discrete Algorithms, 745–754.

International Journal of Computer Applications Technology and Research
Volume 5– Issue 6, 358 - 363, 2016, ISSN:- 2319–8656

Features Analysis in CBIR Systems

Poonam Bhatia Anand
MMICT&BM
Maharishi Markandeshwar University
Mullana, Ambala
India

Neetu Bhatia
MMICT&BM
Maharishi Markandeshwar University
Mullana, Ambala
India

Abstract: Image retrieval is the major innovations in the development of images. Mining of images is used to mine latest information from the general collection of images. CBIR is the latest method in which our target images is to be extracted on the basis of specific features of the specified image. The image can be retrieved in fast if it is clustered in an accurate and structured manner. In this paper, we have the combined the theories of CBIR and analysis of features of CBIR systems.

Keywords: Image retrieval, Content based image retrieval, CBIR techniques, Systems and features.

1. INTRODUCTION

In the present era, images play a big role in every part of life. As images are increasing day by day on the web, Retrieval of efficient and accurate images from the web is very crucial task in the field of image retrieval. Presently there are two methods that are used for retrieving the images. One method is Text Based Image Retrieval in which retrieval of images is done using textual features of the images. This method does not use the visual features of the images. Due to this drawback some problems exist to meet accurate images in a right way [19] when we search the image database. On the Other Hand, the second method used for retrieval of image is CBIR. Its goal is to search the image from a huge image database according to the query given by the user. In CBIR, the color, texture and shape are visual features extracted from the image. Therefore it is also called visual image Retrieval. In the next section the model of CBIR is discussed.

2. CONTENT BASED IMAGE RETRIEVAL

CBIR uses the exclusive part of image to signify and true to use. In common CBIR system is divided into following three phases which are shown in figure 1:

- Extraction of features
- Feature Matching
- Semantic image retrieval

In this system, first the query to the system provide by the user. The user's query can be either in the form of wording or in the form of an picture. Using feature extraction techniques, from the image, the features like color, texture and shape are extracted. After that, these features are matched with the images that are stored in the database of images. Then various similarity measures are used by a variety of systems to allocate the index value of the images or objects.

CBIR is considered as one of the efficient ways for retrieving images. It points at budding new procedure that supports well-organized probing and browsing of huge libraries of digital image depend on derived features of the images. First, the image is divided into chunks of equal size. This system uses the image content directly that will be searched in an image database. The key plan of this is to analyze information of the image by using features of color, texture, shape, face detection and layout of an image.

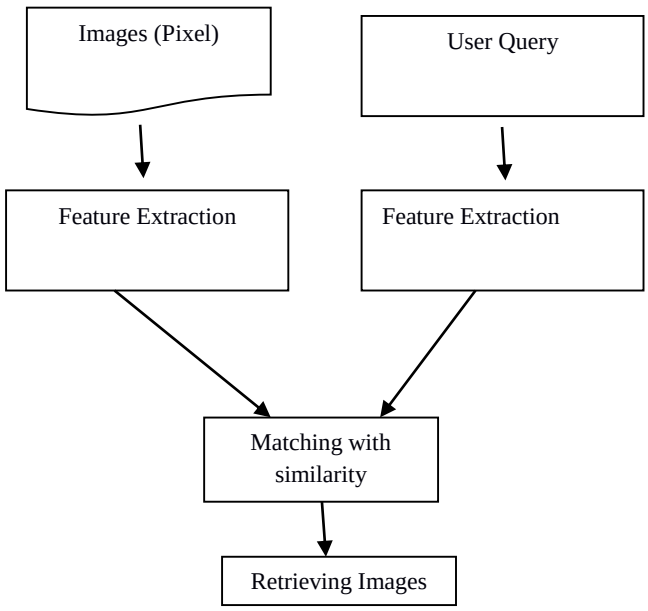


Fig 1: CBIR System [21]

The various features used in CBIR Systems are explained in the next section.

3. FEATURES USED IN CBIR

Features are the properties of the image. It describes the image in various ways. There are two types of features used in CBIR Systems i.e. Low Level Features and High Level Features [3,12]. These Features are described as follows:

3.1 Low level features

The primary features used in CBIR Systems are low level features. Following are the basic low level features:

3.1.1 Color

This feature refers to the color of different parts of the images. This feature can be extracted by methods such as Histogram method Arithmetical method Color / Shade model.

3.1.2 Texture

Texture refers to visual patterns in images. Texture is used to represent texture in images. Texture is also one of the important kind of an image. This feature refer to the visual patterns including the surface of clouds, trees, bricks, hair, and fabric. A variety of algorithms have been proposed for texture testing: Gray Level Co occurrence, The Tamura Texture Feature, The Model of Markov Random Field, Gabor Filtering, Binary Local Patterns.

3.1.3 Shape

Shape refers to shape of particular region in the image. Shapes can be determined by applying segmentation of the image. Shape Descriptors are used to translate, rotate and scale the images. Boundary based and Region based Shape Representation is used to represent the image.

3.2 High level Features

The high level features are discussed as below:

3.2.1 Metadata [19]

Meta data is data about data that are related to the process of the image creation. The Meta attributes include image acquisition date, image identification number and name, image modality device, image magnification, etc. Due to manual annotation in images retrieval following limitations exists:

1. Excess time consumption
2. Expensive task for large image database
3. Retrieval of non subjective, context sensitive and incomplete data.

Therefore CBIR is used for image retrieval to overcome the disadvantages of keyword annotation method.

3.2.2 Semantics

Semantic image retrieval [7] begins by request made by a person, e.g. "find pictures of Mahatma Gandhi". The above said task is very difficult to perform for computers as you will not find Mahatma Gandhi always posing in the same pose or in front of the camera. To evade this problem, CBIR systems uses feature of low level such as color, shape and texture. The combination of these features is used with databases which are already well equipped and qualified to match these features, i.e. fingerprints, faces or shapes, etc. Although in general, high-level identification concepts of image retrieval require human feedback.

In the next section CBIR Systems are explained.

4. CBIR SYSTEMS

There are lots of areas such as academia, commerce, government and hospitals in which large group of digital images are being formed. Some of those images are the manufactured goods of digitizing accessible group of diagrams such as drawings, paintings, and prints in these image collections [8]. Normally, keyword indexing is the only method of searching this collection. In this section we explain some technical phase of current CBIR systems. There are number of image database systems, retrieval of images, or Systems of multimedia information has been published. The major target of this clarification is to provide a review of the short-term CBIR systems.

Some of the commonly used CBIR Systems are described as follows:

4.1 Alexandria Digital Library(ADL)

In this CBIR system [17], the only texture feature is extracted from the images using with texture features images can explore. With the help of browser map the user can interrelate magnifies a two dimensional world map to find its significance area, and choose a question area that has to involve with the database images. Right now, a list of images, query parameters can be selected by the user for example, aerial of the photo, map, images of remote sensor etc. and then with the help of browser map the images that are overlapped with area are retrieved.

4.2 Advanced Multimedia Oriented Retrieval Engine(AMORE)

In this CBIR system [18], the images are divided into eight sections by region or all the same color and downsize into pixels of 24*24. At firs, the user selects a different group of images. The first image group is being chosen at random. Then selected images, generally

parallel descriptions images can be getting back. The query of images can also be specific by its Uniform Resource Locater. Then the comparative significance of color and shape is indicated by user.

4.3 Blobworld

In this CBIR system [5], the color, shape of regions which are called blobs, texture and location are features of the query image. In Lab space the color description is done by the coordinates of the histogram of 218 bins. First, categories have the limiting search is selected by the user. After selecting a blob, the user points to the importance of the region such as somewhat and very. After then the user points at blob's color, texture, location, and shape. For querying, more than one region can be used by the user.

4.4 Candid

In this CBIR system [13], each image is represented by signature containing Gaussian functions. The clusters of spectral bands are used by color features. The meaning of every pixel and Skegness are used by texture. A query image is provided by the user.

4.5 Content based Visual Query

The system enables queries by texture is called as CBVQ. First, a 9-dimensional texture vector is calculated for each pixel in the image. After then by using the filtering of non linear techniques, the pixels of the image are merged into texture regions of homogeneous to reduce noise followed by algorithm of a sequential labelling then by overlapping regions and renovation of the images is done from the images of the binary sub band [23]. By using 9 dimensional binary vectors, each region is represented. Spatial information is also mined for each texture region detected; a global color histogram is also calculated for every image database. With queries by example the system allows outline in which one of the display random images can be selected by the user or address of any image gives the URL and by a color histogram user can give direct queries. The available search methods is chosen by the user selects in a query for example histogram of color, manipulation of texture To build a new query histogram, a histogram of the query image can be used.

4.6 Chabot

This system is made up of an energetic text record of the collection, the location of the picture[20]. A 20 bins of color histogram are worked out for each image. There is certeria of a search are presented by user such as keywords, location and colors. Although there are some partial options of the color criterion for the user. The user has the many options to organize thoughts, which is combination of criteria of search in which thought make happy. For instance, as a sunset is grouping of keyword and mainly red or mainly orange is the color criterion used for the concept of 'sunset' is described.

4.7 Flexible Image Database System(FIDS)

In this CBIR system [4], the features are also taken in the formation of the sub image grid such as rows and columns. The histogram values is obtained after applying the filter of Sobel edge. The inquiry image as an image is chosen by the user. The feature distance measures can also be chosen by the user, and merge them. As a next inquiry image the result image is displayed.

4.8 Fast Object Color Based Query System(FOCUS)

In this CBIR system [6], every image is distributed in a group of cells having dimensions of 100×100 pixels. In HSV space each cell is grouped as a color histogram. SPG (Spatial proximity graph) is built in many stages to obtain the target of characterizing spatial relationships between color regions. The base of connecting two nodes is

a) In the same cell resultant peaks situated.
b) Resultant peaks situated in different groups, but have the same color.

All linked nodes of the same color are then unified by merging them to a single node. Adjacency matrix representation is then used for storing linked nodes. A global color histogram is calculated and determined the relationship of color region for the inquiry image at the pixel level.

4.9 Jacob

The color and texture is used in this CBIR system and a histogram in RGB space is used for Color characterizing [15]. In this, query may be straight. By inserting a few standard queries which made straight by indicating the histogram and/or features of the texture. The user must give an image, for an example of a query.

4.10 Metaseek

In this CBIR system [3], to select target searching engines in a presentation, list MetaSeek is used. After then QBIC, VIRAGE, WebSeek and the VisualSeek are the CBIR systems followed for the real matching. The client can choose a class and offer a keyword, then give a URL of an image. With the help of Metaseek, the sender validates the list presentation performance upon acceptance of a query. These databases contain scores of history query successes.

4.11 Multimodel Information Retrieval System

This system merges various techniques to get images of descriptions semantically. With the help of natural language processing (NLP) techniques, picture's information is extracted in analyzing image captions [24]. This information consists the people with their names, location and time of picture taken with optical description such as gender and hair color. The pictures are detected by NLP in which peoples are presented which are identified by face to verify the guess. When the area of face's is cropped out then for remaining images, a histogram is calculated. For the scenery, a histogram of color correlation is computed in which no face is identfied by the NLP analysis.

By consisting of text string, an image and different topics about sports, politics and entertainment, query is constructed. At this, the relative significance of text vercus content as well as foreground versus background can also be painted by the user.

4.12 Netra

In this CBIR system [16], By using color, the database of Images is created by division of images. To characterize the shape of regions, three feature vectors are used. The first one is the curvature function of the curve. The second one is the region of the centroid function and the third one is the function of complex coordinate. In 25 categories, there are 2500 Corel photo images which are gathered and organized with 100 images in each category. As the enquiry image, user can decide on any one of them. All database images are segmented in all the same regions. Color, spatial location, texture and shape are main characteristics of the image which the user can choose. Color and spatial location can be directly used for image example. The spatial location is used to illustrate the area of interest of two bound box, for confine the favorite region point to the inner box and for confine the region of objects point to the outer box.

4.13 Photobook

In this CBIR system [22] , three types of looms are applied to create an image demonstration for purpose query, each for a specific type of image contents: faces, 2D shapes and images of texture. The first two demonstrations are same because they propose an explanation by using covariance of the eigenvectors. To perform a query for present an explanation filter is used from network of images the user chooses from a little explanation of images. From that display, another query image might be selected by the client and repeats the search.

4.14 Pictoseek

The queries based on content color and shape are classified as features in this CBIR System [10]. Different color, features, the effect of various imaging conditions is examined below the statement of a model of dichromatic reflection. In this system reflection is used for designing that how a cluster of color is constructed. By the system's Web crawlers, with help of giving a URL address the query images from the database is chosen to compose. The desired variance is chosen by the user before giving the query, which points to the system.

4.15 Query by Image Content (QBIC)

In this CBIR system [26] the features of texture are used with customized versions of the roughness, dissimilarity. We put queriea by using image content. The shape area is used by shape features circularity, eccentricity, major axis orientation. These shape features are planned partially involuntarily and mined for all objects. On the base of image example, color, patterns and constructed sketch by user, queries are allowed by QBIC system. By using the slider, the percentage of color is adjusted in the image.

4.16 Virage

This system is an extensible framework structure for constructing CBIR systems [1]. In this CBIR system, the basic concept is the type of feature, computation and distant matching. There are five abstract data types: a) global value, b) color histograms, c) local value, d) histograms and e) graphs. The VIRAGE system gives a set of general primitives. The VIRAGE system gives a GUI tool set essential for the growth of a GUI edge. These contain services for placing of images, image queries and maintain many popular formats of image file. Query by sketch is also used where picture can be sketch by user with drawing tools and color palette.

4.17 Visualseek

In this CBIR system [23], using the back-projection technique, the color region extraction takes place. In this system of the population, every image is decayed robotically into region of foremost colors. Features and Spatial properties are used in every region. For start a query, a number of regional sketches by the user and selects a color for each region. After then it points to location boundary, size and relationship between regions.

5. ANALYSIS OF FEATURES IN CBIR SYSTEMS

Low level features, High level features and Keywords which are used by CBIR Systems are shown in table 1. Some systems uses high level features and some systems used low level features and some systems uses keywords. High level features are face detection and layouts. Color, shapes and texture are low level features. Color features can be described as Color moments, Global histogram, Dominant colors and Correlation histogram. Shape features are described as

Table1: Feature Analysis in CBIR Systems

CBIR SYSTEMS	Features														
	Low Level Features												High Level Features		
	SHAPE				TEXTURE				COLOR						
	Fourier Descriptor	Elementary Descriptor	Bounding Box	Template Matching	Atomic Texture	Wavelet, Gabor, Fourier	Edge Statistics	Random Fields	Dominant Colours	Colour Moments	Global Histogram	Correlation Histogram	Face Detection	Layout	KEY-WORDS
ADL						Yes									Yes
AMORE				Yes					Yes						Yes
BLOBWORLD		Yes			Yes									Yes	
CANDID					Yes				Yes						
CBVQ		Yes				Yes					Yes				
CHABOT											Yes				Yes
FIDS						Yes	Yes				Yes			Yes	
FOCUS		Yes												Yes	
JACOB					Yes						Yes	Yes			
METASEEK					Yes						Yes				Yes
MIR					Yes	Yes					Yes		Yes	Yes	Yes
NETRA	Yes					Yes								Yes	
PHOTOBOOK								Yes					Yes		
PIC TO SEEK	Yes									Yes	Yes				
QBIC		Yes		Yes	Yes					Yes	Yes			Yes	Yes
VIRAGE															Yes
VISUALSEEK		Yes	Yes						Yes					Yes	

Fourier descriptor, Elementary descriptor, Bounding box and Template matching. Texture features are described as Atomic texture features, wavelet, gabor, fourier edge statistics and random fields etc.The eigen imae is an example of dominant color, because it's feature is copied from the value of global image color. The features of atomic texture are represents gap of an image, regularity, directionality and smoothness of an image. The edge statistics represents histogram of image and orientation. The centroid, quarter, orientation axix, strangeness are the kinds of elementary shape descriptors. The layout feature represents the total or partial locality of the color, shape and texture.

The simple features can be easily mined, and it is easy to apply in the CBIR systems, and the easiest to use in the accurate method At last, color feature is established effective, for the reason that good quality color feature is not very complicated to design and implement. As retrieval of images is very challenging task so .

Researchers have proposed and used various CBIR Systems according to various characteristics of systems. These CBIR Systems perform the task of retrieving the images based on the various features available in it. Some systems retrieve the images by using only color, texture or shape and some uses only keywords or metadata to retrieve the images there are very less number of systems which uses both low level and high level features to retrieve the images in effective manner.

From table 1 it can be shown that QBIC is the most effective system that is widely used by the researchers and if we talk about the features in this system then it used all the features like color, texture,shape, layout as well as keyword based retrieval of images. According to the features used by various CBIR Systems, QBIC uses both high level features, low level features as well as keywords for retrieving images according to the requirements of user.

6. CONCLUSION

The techniques of CBIR are still under beneath investigation. To increase the performance of image retrieval capability of systems of CBIR are used. In this paper we have discussed many features used in CBIR systems. It can be drawn from the analysis that the wide varieties of features are used by the QBIC CBIR systems. Still this area needs more research for finding a best system for image retrieval process so that images could be retrieved in effective manner. Most systems are under research is the main product of these systems.

7. REFERENCES

[1] Bach J.R., Fuller C., Gupta A., Hampapur A., Horowitz B., Humphrey R., Jain R.C. and Shu C.F., 1996. Virage image search engine: an open framework for image management. In Electronic Imaging: Science & Technology (pp. 76-87). International Society for Optics and Photonics.

[2] Balan S., and T. Devi. 2012. Design and Development of an Algorithm for Image Clustering In Textile Image Retrieval Using Color Descriptors." International Journal of Computer Science, Engineering and Applications (IJCSEA) Vol. 2(3).

[3] Benitez Ana B., Mandis Beigi, and Shih-Fu Chang. 1998. Using relevance feedback in content-based image metasearch." Internet Computing, IEEE vol. 2(4) ,pp.59-69.

[4] Berman Andrew P. and Linda G. Shapiro.1998. A flexible image database system for content-based retrieval. Pattern Recognition,. Proceedings. Fourteenth International Conference on. Vol. 1. IEEE.

[5] Carson, C., Thomas, M., Belongie, S., Hellerstein, J. M., & Malik, J. 1999. Blobworld: A system for region-based image indexing and retrieval. Visual Information and Information Systems. Springer Berlin Heidelberg.

[6] Cerra D. & Datcu, M. 2012. A fast compression-based similarity measure with applications to content-based image retrieval. Journal of Visual Communication and Image Representation, vol. 23(2), pp. 293-302.

[7] Chadha A., Mallik S. and Johar, R. 2012. Comparative study and optimization of feature-extraction techniques for content based image retrieval, arXiv preprint arXiv:1208.633.

[8] Chadha D. and Singh N. 2012. Query by Image for efficient Information Retrieval- A Necessity, International Journal of Computer Applications.

[9] Chary R., Lakshmi, D. R. & Sunitha, K. V. N. ,2012. Feature extraction methods for color image similarity. arXiv preprint arXiv:1204.2336.

[10] Gevers T., & Smeulders A. W., 2000. PicToSeek: combining color and shape invariant features for image retrieval. Image Processing, IEEE Transactions on, vol. 9(1), pp.102-119.

[11] Huang T. Rui, Y., & Chang, S. F. 1997. Image retrieval: Past, present, and future ,In International Symposium on Multimedia Information Processing (Vol. 108).

[12] Kalel, M. D. S., Pisal, M. P. M., & Bagawade, M. R. P.: 2016. Color, Shape and Texture feature extraction for Content Based Image Retrieval System: A Study, International Journal of Advanced Research in Computer and Communication Engineering, vol. 5(4).

[13] Kelly P. M., Cannon, T. M., & Hush, D. R. 1999. Query by image example: the comparison algorithm for navigating digital image databases (CANDID) approach. In IS&T/SPIE's Symposium on Electronic Imaging: Science & Technology (pp. 238-248). International Society for Optics and Photonics, March 1

[14] Khokher A. & Talwar R. 2011. Content Based Image Retrieval: State-of-the-Art and Challenges. International Journal of Advanced Engineering Sciences and Technologies, vol. 9(2), pp. 207-211.

[15] La Cascia M. and Ardizzone, E. 1996. JACOB: Just a content-based query system for video databases" In Acoustics, Speech, and Signal Processing, 1996. ICASSP-96. Conference Proceedings. IEEE International Conference on IEEE., Vol. 2, pp. 1216-1219.

[16] Ma W. Y. and Manjunath B. S. 1999.Netra: A toolbox for navigating large image databases. Multimedia Systems,vol. 7(3),pp. 184–198.

[17] Manjunath B. S. 1995. Image Browsing in the Alexandria Digital Library (ADL) Project D-Lib Magazine, vol. 1(2).

[18] Mukherjea S., Hirata, K., & Hara, Y. 1999. Amore: A world wide web image retrieval engine. World Wide Web, vol. 2(3), pp. 115-132.

[19] Nalini M. P. & Malleswari B. L. 2016. Review on Content Based Image Retrieval: From Its Origin to the New Age., International Journal of Research Studies in Science, Engineering and Technology Vol. 3(2), pp. 18-41.

[20] Ogle V. E., & Stonebraker, M. 1995.Chabot: Retrieval from a relational database of images. Computer, vol. 28(9),pp. 40-48.

[21] Pattanaik, S. Prof. DG Bhalke. Beginners to Content Based Image Retrieval. International Journal of Scientific Research Engineering & Technology, 1, 040-044.

[22] Pentland A., Picard, R. W., & Sclaroff, S. 1996. Photobook: Content-based manipulation of image databases. International journal of computer vision, vol. 18(3), pp. 233-254.

[23] Smith J. R. & Chang, S. F. 1997. Querying by color regions using the VisualSEEk content-based visual query system. Intelligent multimedia information retrieval, vol. 7(3), pp. 23-41.

[24] Srihari R. K., Zhang, Z., & Rao, A. 2000. Intelligent indexing and semantic retrieval of multimodal documents". Information Retrieval, vol. 2(2-3), pp. 245-275.

[25] Su, J. H., Huang W. J., Yu P. S. & Tseng, V. S. 2011. Efficient relevance feedback for content-based image retrieval by mining user navigation patterns. Knowledge and Data Engineering, IEEE Transactions on, vol. 23(3), pp. 360-372.

[26] Wayne, N., Ron, B., Will, E., Flickner, M. D., Glasman, E. H., Petkovic, D.,& Taubin, G. 1993. The QBIC project: Querying images by content using color, texture, and shape. In Proc. SPIE ,Vol. 1908, No. 1, pp. 173.

International Journal of Computer Applications Technology and Research
Volume 5– Issue 6, 364 - 367, 2016, ISSN:- 2319–8656

A Review of Storage Specific Solutions for Providing Quality of Service in Storage Area Networks

Joseph Kithinji
Department Computer Science and Information Technology,
School of Information Technology and Engineering,
Meru University of Science and Technology,
Meru, Kenya

Abstract: Predictable storage performance is a vital requirement for promising performance of the applications utilizing it and it is the systems administrators' job to ensure that storage performance meets the requirements of the applications. Most storage solutions are able to virtualize the amount of storage presented to the host in a flexible way, but the same storage devices have no QOS features .Storage level agreements provided by storage devices do not provide predictability in service delivery due to the absence of prioritization (QOS) mechanisms in storage devices. This paper reviews some of the storage specific solutions developed to implement quality of service in storage area networks.

Keywords: Starvation, latency, burst handling, quanta, performance isolation.

1. INTRODUCTION

Storage area networks play a key role in business continuity, enterprise wide storage consolidation and disaster recovery strategies in which storage resources are most often distributed over many distant data centers[10].Future storage systems are required to scale to large sizes due to the amount of information that is being generated. In a SAN large numbers of magnetic disks are attached to a network through custom storage controllers or general purpose pcs and provide storage to application servers [6].

In a SAN, a single host request may flood the resources of a storage pool causing poor performance of all hosts utilizing that particular pool [5]. Hence, the performance of a given host utilizing a shared pool resource is unpredictable by the nature of resource sharing. To address this problem a mechanism of providing QOS based on some policy is required. Storage service level agreements provide for predictability in service delivery which is not effective due to the absence of QOS mechanisms in storage devices [5].

QOS is essential in the mixed environment where various users with different levels of priorities and preferences are accessing the storage systems simultaneously. For example in an enterprise network, web hosting, data analysis and data editing may be running at the same time[10].Providing QOS to SANs has been a challenge which has led to the design of many approaches.

2. STORAGE SPECIFIC QUALITY OF SERVICE SOLUTIONS

2.1 Stonehenge

QOS is essential in mixed environment where various users with different levels of priorities and preferences are accessing the storage systems simultaneously.it is important to ensure that critical tasks get satisfying performance given limited resources.[8] Developed Stonehenge to solve the issues of storage scalability, manageability and quality of service. Stonehenge is built on IP networks IDE hard drives, IDE controllers and off-the shelf low end personal computers. To implement QOS Stonehenge dedicates a set of storage servers to manage disk arrays and single personal computers to perform the controlling functions such as storage reservation and run-time management [7].

[1]Developed pClock based on arrival curves that capture the bandwidth and burst requirements of applications. When implemented pClock showed efficient performance isolation and burst handling.it also showed an ability to allocate spare capacity to either speed up some applications. When a request arrives the pClock algorithm performs three functions; updating the number of tokens, checking and adjusting tags and computing the tags. The update number of tokens function updates the arrival upper bound function for the present arrival time while the check adjust tags is used to-resynchronize flows to avoid starvation and the compute tags assigns start and finish tags. The pClock algorithm allows multiple workloads to share storage, with each workload receiving the level of service it requires. PClock allows each workload to specify its throughput, burst size and desired latency [1] [8].

The pClock algorithm is as follows:-

Request arrival:

Let t be arrival time of request r from fi;

Update Numtokens();

CheckandAdjustTags();

ComputeTags();

Request scheduling:

Choose the request w with minimum finish tag *fjw* and dispatch to the server

Let the chosen request be from flow fk with start tag swk;

Minsk=sk; [5].

2.2.1 updateTokens

In order to assign tags the arrival upper bound function Uia() to the current time t.it maintains a variable numtokens for each flow fi.

This means that the difference Uia(t)-Ri(0,t) is the difference between AUB at time t and the cumulative number of arrivals up to that time. The value obtained indicates the number of requests that can be made by fi at t without violating the arrival constraints.

Hence when [9];

Uia(t)-Ri(0,t)<1, means that a well behaved flow cannot make any request at t.

2.2.2 computeTags

This function assigns start and finish tags (Sir + Fir) to the request r from fi arriving at time t.The value assigned to the start tag Sir depends on whether the request is within the AUB or exceeds it.When numtokensi>=1, Sir is set to the current time t.If the total number of requests made by fi through time t exceeds AUB(numtokensi<1),the start tag will be assigned a future time greater than t.In particular the start tag is set to the time it would have taken a well behaved flow to send a number of requests[2].

Pclock guarantees that the well behaved flows are not missed and the requests of the background jobs are done in batches, which can lead to better disk utilization since many background jobs tend to be sequential [8]. The algorithm has the ability to allocate spare capacity to the workloads or to the background jobs. The algorithm is also lightweight to implement and efficient to execute. However it does not offer control of how QOS mechanisms interact with storage devices [7].

2.3 Argon

The argon storage server explicitly manages its resources to bind the inefficiency arising from interservice disk and cache interference in traditional systems. The goal is to provide each service with at least a configured fraction of the throughput it achieves when the storage server to itself. Argon uses automatically configured prefetch/write back sizes to insulate streaming efficiency from disk seeks introduced by competing workloads. Argon uses prefetching and write back aggregation as a tool for performance insulation [4] [6].

Argon adapts, extends and applies some existing mechanisms to provide performance insulation for shared storage servers. Many operating systems such as eclipse operating system use time slicing of disk head time to achieve performance insulation. Argon goes beyond this approach by automatically determining the lengths of time slices required and by adding appropriate and automatically configured cache partitioning and prefetch/write back [8].

Argon uses QOS aware disk scheduler in place of strict time slicing, for workloads whose access patterns would not interfere when combined.to implement fairness or weighted fair sharing between workloads argon uses amortization cache partitioning and quanta based scheduling. Assumes that network bandwidth and CPU time has no effect on efficiency [9]. To achieve complete isolation argon does not allow requests from different workloads to be mixed, instead it uses a strict quanta based scheduling .This ensures that each client gets exclusive access to the disk during a scheduling quantum which avoids starvation because active clients quanta are scheduled in a round robin manner[5].

Traditional disk and cache management allow interference among services access patterns to significantly reduce efficiency [7]. Argon combines and automatically configures prefetch/write back cache partitioning and quanta based disk time scheduling to provide each service with a configurable fraction of efficiency it would receive without competition. This increases both efficiency and predictability when services share storage server [4]

However as with all other storage specific solutions Argon runs on the storage device itself which requires multiple instances of it to be implemented in all the devices. This increases overhead and CPU time. Again since there is no centralized management of QOS when the storage data is in transit from the source to destination QOS is not taken care of. The argon design also assumes that bandwidth is not a factor in QOS however with IP SANs bandwidth management is very important since the storage data will be moving from source to destination via IP network [6].

2.4 Facade

[3]Developed Façade as a dynamic storage controller for controlling multiple input/output streams going to a shared storage device and to ensure that each of the input/output streams receives a performance specified by its service level objective. Façade provides performance guarantees in highly volatile scenario. To achieve QOS Façade is implemented as a virtual store controller that sits between hosts and storage devices in the network, and throttles individual input/output requests from multiple clients so that devices do not saturate [2].

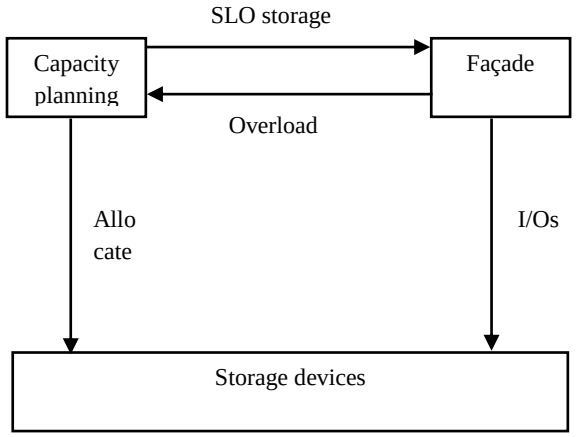


Figure: Facade structure [3]

The capacity planner allocates storage for each workload on the storage device and ensures that the device has adequate

capacity and bandwidth to meet the aggregate demands of the workloads assigned to it.The allocation is adjusted depending on the workload. Requests arriving at façade are queued in per workload input queues.to determine which requests are admitted to the storage devices façade relies on three components that is the I/O scheduler, statistics monitor and controller [8].

The I/O scheduler maintains a target queue depth value and per workload latency target which it tries to meet using earliest deadline first (EDF) scheduling. The deadline for a request from a workload WK is arrivalTime (WK) + latenctyTarget (Wk), where arrivalTime (WK) is its arrivalTime and latencyTarget (WK) is a target supplied for WK by the controller. Requests are admitted into the devices in two cases; if the device queue depth is now less than the current queue length target or if the deadline for any workload is already past. The intent of controlling queue depth is to allow workloads with low latency requirements to satisfy their SLOs [3].

2.4.1 Statistics monitor

This receives I/O arrivals and completions.it reports the completions to the I/O scheduler and also computes the average latency and read and write request arrival rates for active workloads every P seconds and reports them to the controller [10].

2.4.2 Controller

The controller adjusts the target workload latencies and the target device queue length. Target workload latencies must be adjusted because the workload request rates vary and therefore it is necessary to give those requests a different latency based on the workload SLO.The device queue depth must also be adjusted to meet the varying workload requirements[8] .The controller tries to keep the queue as full as possible to enhance device utilization. However this increases the latency. This means when any.

Workload demands a low latency, the controller reduces the target queue depth. The controller uses the I/O statistics it receives from the monitor every P seconds to compute a new latency target based on the SLO for each workload as follows;

Let the SLO for WK be ((r,tr1,tw1),(r2,tr2,tw2),…,(rn,trn,twn)) with a window w and the fraction of reads reported is fr.

Let r0=0, rn+1=∞, trn+1=twn+1=∞ then latencyTarget (WK) =trifn+twi (1-fr)

If ri-1<=readRate (WK) + writeRate (WK) <ri [7].

Facade is able to efficiently utilize resources and balance the load among multiple backend devices while satisfying the performance requirement of many different client applications. Facade is also able to adopt to workloads whose performance requirements change overtime. However façade cannot handle large workloads. This is because multiple instance of façade that are in every storage device cannot be able to cooperate in order to handle large workloads [3].

2.5 PARDA

PARDA enforces proportional share fairness among distributed hosts accessing a storage array without assuming any support from the array itself.PARDA uses latency measurements to detect overload and adjust issue queue lengths to provide fairness [7]. Numerous algorithms for network QOS have been proposed, including many variants of fair queueing.However these approaches are suitable only in centralized setting where a single controller manages all requests for resources [2].

3. Discussion and Conclusion

PARDA enforces proportional fairness among distributed hosts accessing a storage array, without assuming any support from the array itself. PARDA uses latency measurements to detect overload and adjust issue queue lengths to provide fairness. However these technique require each storage device to run an instance of the algorithm which results in overhead caused by running the algorithm. Facade provides performance guarantees by throttling individual input/output requests from multiple clients so that devices do not saturate. Facade provides performance isolation in that the performance experienced by the workload from a given customer must not suffer because of variations on the workloads from other customers. Façade is able to use resources much more efficiently and to balance the load among multiple back end devices while satisfying the performance requirements of many different client applications. However it cannot handle well large workloads. It also requires multiple instances of the same algorithm to run in all storage devices.

Stonehenge was developed as a technique for providing QOS guarantees in storage area networks. All the above techniques require that multiple instances of the same algorithm runs on every storage device. These increases overhead which is caused by the processing of the algorithms. These techniques are implemented on the storage device and therefore so do not provide service guarantees when storage traffic is traversing the network which is important since with IP SAN storage traffic will be interacting with other traffic in the network.

4. ACKNOWLEDGMENTS

My thanks goes to all my friends who have contributed towards development of the paper.

5. REFERENCES

[1] Gulati, A., Merchant, A., & Varman, P. J. (2007). p Clock : An Arrival Curve Based Approach For QoS Guarantees In Shared Storage Systems. ACM.

[2] Gulati, A., & Waldspurger, C. A. (2007). PARDA : Proportional Allocation of Resources for Distributed Storage Access. In 7th USENIX Conference on File and Storage Technologie (pp. 85–98).

[3] Lumb, C. R., Merchant, A., & Alvarez, G. A. (2003). Fac , ade : virtual storage devices with performance guarantees. In File and Storage Technologies (pp. 131–144).

[4] Wachs, M., Abd-el-malek, M., Thereska, E., & Ganger, G. R. (2007). Argon : performance insulation for shared storage servers. In File and Storage Technologies (pp. 61–76).

[5] A. Gulati and I. Ahmad(2008). Towards distributed storage resource management using flow control. ACM SIGOPS Operating Systems Review, 42(6):pp 10–16.

[6] Bjørgeengen, J. (2010). Using TCP / IP traffic shaping to achieve iSCSI service predictability. In Proceedings of LISA '10: 24th Large Installation System Administration Conference (pp. 91–107).

[7] Traver, L, Tarin, C, Cardona, N.(2009)Bandwidth Resource Management for Neural Signal Telemetry, Information Technology in Biomedicine, IEEE , Vol.13, no.6(December), pp.1083-1084.

[8] M.Wachs, M. Abd-El-Malek, E. Thereska, and G.R. Ganger (2007). Argon: performance insulation for shared storage servers. In Proceedings of the 5th USENIX conference on File and Storage Technologies,pages 5–5. USENIX Association.

[9] Van der Stok,P, D. Jarnikov,D,Kozlov, S,van Hartskamp,M & Lukkien,J.J(2009)Hierarchical Resource Allocation for Robust In-Home Video Streaming, Journal of Systems and Software.Vol. 80, no. 7(February), pp. 951–961.

[10] Peter, M.O, and Babatunde, P.J. (2012) Software Prototyping: A Strategy to Use When User Lacks Data Processing Experience, ARPN Journal of Systems and Software.Vol.2,No.6(June),pp 219-223

Table1.Comparison of storage specific quality of service solutions

	QOS measures				
QOS solutions	**Performance isolation**	**Burst handling**	**Bandwidth sharing**	**Centralized control of QOS**	**Control latency**
Argon	✓	×	✓	×	✓
Façade	✓	×	✓	×	✓
Stonehenge	✓	×	✓	×	✓
PARDA	✓	✓	✓	×	✓
pClock	✓	✓	✓	×	✓

International Journal of Computer Applications Technology and Research
Volume 5– Issue 6, 368 - 375, 2016, ISSN:- 2319–8656

Providing a multi-objective scheduling tasks by Using PSO algorithm for cost management in cloud computing

Masrour Dowlatabadi,
Young Researchers and Elite
Club, Sabzevar
Branch, Islamic Azad University,
Sabzevar, Iran

Arman Kavosi Ghafi
Department of Computer
Software, Central Islamshahr
Branch, Islamic Azad
University, Iran

Rasoul Habibzadeh Shojaee,
Department of Computer
Software, Central Islamshahr
Branch, Islamic Azad
University, Iran

Abdolreza Andalib
Department of Computer
Software, Qeshm international
Branch, Islamic Azad
University, Qeshm, Iran

Abstract:
This article is intended to use the multi-PSO algorithm for scheduling tasks for cost management in cloud computing. This means that any migration costs due to supply failure consider as a one objective and each task is a little particle and recognize by use of the appropriate fitness schedule function (how the particles arrangement) that cost at least amount of total expense. In addition to, the weight is granted to the each expenditure that reflects the importance of cost. The data which is used to simulate proposed method are series of academic and research data that are prepared from the Internet and MATLAB software is used for simulation. We simulate two issues, in the first issue, consider four task by four vehicles and divide tasks. In the second issue, make the issue more complicated and consider six tasks by four vehicles. We write PSO's output for each two issues of various iterations. Finally, the particles dispersion and as well as the output of the cost function were computed for each part.

Keywords: Cloud computing, Optimization particle algorithm, Apportion duties, MATLAB software

1. INTRODUCTION

The basic concept of cloud computing dates back to 1995, When John McCarthy stated: Cloud computing may someday be organized as a public industry [1]. Cloud computing, briefly Consist of required computing, provide required fields and presenting computer services by a cloud of software, hardware and IT resources [2] [3]. Usually diagram of computer networks, the figure of the cloud is used as a summery way to hide the complex infrastructures that are inside the network. Cloud word in the cloud computing is extracted of these metaphor and is used to hide the Internet and infrastructures in the cloud computing [4]. By the cloud computing people use the PC to achieve a network of computers that are invisible status [5]. What makes this concept a scalable and efficient is a Service-Oriented Architecture (SOA) [6] [7]. Resources, are kept on servers providers instead of to be installed on a user's system.

By assume cloud computing, develop cloud concept, in addition to servers cover infrastructures of network. In mid-2008, Gartner found that there was an opportunity in cloud computing that was created for « Shaping the relationship between the consumers of IT services, between those who consume these services and those who sell these services ». That said, the cloud computing, will be the foundation in the IT industry for the next 20 years [8].Cloud computing is a new technology, which in recent decades by various companies, most notably Microsoft, Google and Amazon have been released to the world. Cloud computing technology is a great and safe storage space, which is always different with processing and computing power. In this regard, different companies, offered different services with various computing services [9]. These calculations are based on TCP / IP (internet-based) and include processors, massive memories, quick data transfer networks and architecture of reliable systems and without the governing standard network protocols, cannot be brought the entity to the technology [10]. Different types and levels of cloud services mean it is important that, we are working with which cloud computing systems [11]. A workflow, is an automation of processes to work on the documents, information, and tasks that passes from one process to another and usually is modeled in the form of a circular oriented graph (DOG) [12]. An example of this oriented graph can be seen in Figure 1. So that, each node shows a task and each edge shows dependencies between functional application tasks. Simply, the Process of mapping tasks in a workflow computing resources to perform tasks (To maintain dependencies between them) is defined workflows scheduling. The correct scheduling can have a great effect on system performance. The mapping of tasks to computing resources belonging to the class of things that are well-known NP-Complete [12] [13]. Useful scheduling algorithm can meet user requirements and improve resource efficiency. The technology of cloud computing can be measured like telephone and electricity companies and sharing with customers [14]. Currently, Microsoft's cloud-based service share with customers as Microsoft Online service [15].Cloud computing models resources, similar to a public industry (electricity, telephone, etc.). As without the preparation and storage of expensive tanks, capacitors or transformers. Only when needed, we use electricity and water and only pay based on our consumption. In the case of software and other resources, we just open the software (Eg, tap) and only pay the amount that we consumed. Without need to purchase and maintain expensive resources for their implementation [16]. The billing is based on the amount of minutes or hours that tenants spent on a cloud system, transferred data or filled storage space [17]. Sadri et al (1390) in his study that called "secure message transfer protocol in the cloud computing- based network".

A brief description of a secure messaging system design under internal network according to the cloud computing model with utilizing Rayndl algorithms [18]. Ashoorion (1390) studied messages safety by computing in smart platforms.

He indicated that mobile agents architecture has overcome the many of limitations of architecture on client and server [19]. Delavar Ghorbannia and Arian (1390), provide a combined scheduling method based on genetic algorithm in cloud computing environment [20]. Matin Far (1391) were studied methods, challenges and tools to implement fault tolerance in cloud computing. In this research, the fault tolerant techniques in forms of cloud computing, fault tolerance policies, challenges and the tools used to implement fault tolerance in cloud computing were discussed[21]. Maleki et al (1391) in his study of the CloudSim's simulation model and it's evaluation, introduced CloudSim's simulation tools, that, enables modeling and simulation of cloud computing systems and environments to the preparation of applications [22]. Reza Zadeh (1391) in his study investigated "the tolerability of a bug in the cloud computing".

In these calculations some bug's tolerance models were presented. In these models redundancy has been used to improve the reliability [11]. Shams and Ghiyasi (1392), provided combined scheduling method based on genetic algorithm in cloud computing environment. In this study, they selected the distributed scheduling reliable algorithm in cloud computing (RSDC) due to its simplicity and has been investigated the reliability on the chosen algorithm [23]. Slvarany and Soodha (2010), offered a fee-based algorithms for efficient mapping of tasks to available resources in the cloud. The scheduling algorithm also measures the cost of using the resource and performance of analysis [24]. Raphael et al. (2011) is to minimize power consumption in data centers and improve their weight balance at the same time. In this regard, they stated a scheduling algorithm [25]. Moun and Tou (2011), provided an efficient method for scheduling virtual machine to virtual machine management. The proposed method has been called EVMS. Based on EVMS, scheduler system knows the current status of virtual machines in the cloud [26]. Huang et al. (2012) provided an efficient energy scheduling algorithm (ESS) to reduce energy consumption.

The results show that, ESS has ability to reduce energy consumption significantly by taking SLA's [27]. Huh et al. (2012), provided a task migration algorithm based virtual execution time that limited the maximum difference in virtual execution time between tasks. [28]. Peng et al. (2011), delivered an energy-efficient scheduling method based private cloud. In the paper presented three schedule criteria proposed and provided a combined of energy-efficient scheduling method [29]. The results show that, the efficiency of resources in the enhanced algorithm improved [30].

2. Optimization Particles Algorithm

The swarm particles optimization algorithm is one of the most important algorithms in the field of swarm intelligence. The algorithm introduced by Kennedy and Eberhart in 1995 and by inspiration of the social behavior of animals such as fish and birds live together in small and large groups, have been designed. Other names of this algorithm are the swarm particle algorithm and birds algorithm. PSO, has similarities with evolutionary computation techniques, like genetic algorithms (GA). However, unlike GA, PSO has not evolutionary functions, such as transit and crossover. For successful implementation of PSO, one of the key actions is, discover solution of the problem within the PSO

particles. Correctly affects its performance and its possible implementation (Abraham et al., 2006). Kennedy and Eberhart (1995), demonstrated that PSO of each particle represents a possible answer that moves randomly in problem space. Change the location of each particle in the search space is influenced by its knowledge and neighbors, so the position of the swarm particle affects the way of particle search. As a result of modeling this behavior is the search process in which particles move to the appropriate areas. Particles are taught in groups, based on acquired knowledge, they go to the best of their neighbors. PSO based on the assumption that in every moment of every particle in the search space adjusts its location according to its best location, which has ever been and the best place in the neighboring. The overall design effective implement of PSO is difficult. Also, check out the dynamic environment for real-time applications such as fast convergence feature PSO to follow the target variable (minimum or maximum fitness function) are necessary.

3. The proposed algorithm and simulation

If you want to run your program in parallel and distributed in the cloud. We should know the properties of the environment. As we become more familiar with the characteristics of this environment, we will be able to use it better and more correct and avoid wasting resources. Parallelization of a program can be done at different levels. From the hardware to a single or multiple systems or cluster. Cloud computing architecture has a three-layer architecture which in the infrastructure layers, we are faced with resource management, the base layer is for development of applicable programs on the cloud resources and the software layer provided applications for final users. If we want to improve the management of resources, doing things and scheduling, we must go into the infrastructure layer. If we work on the development of a special issue generally, we should go into the substrate layer, and if we want to focus on how to provide final service for users, we have compiled software layer. Here we focused on the substrate layer, because we are going to develop a way to deal with cloud PSO algorithm. Our Intended use is in the field of scheduling tasks that can be as a final service is available to users and run and optimization is done in the substrate layer.

To be able to run the program in the field of PSO algorithm, it is necessary to provide the substrate layer for programs implement and development. The proposed method evaluation have been carried out by using two tests. we simulate PSO algorithm for the cloud network with 4 and 4 tools .Data by the network as follows:

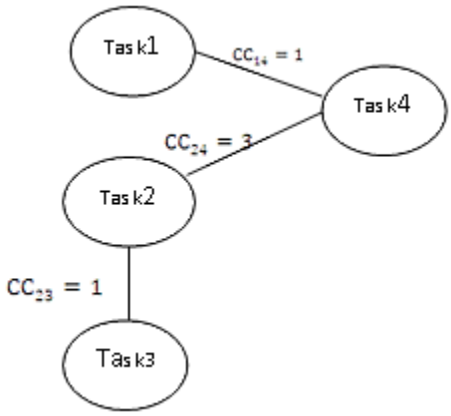


Figure 1. The task dependency graph

Duty as a TDG (task dependency graph) are shown in the following graph. The graph can be converted to the tasks matrix.

Table 1. Tasks Matrix

4	3	2	1	Task#
1	0	0	0	1
3	1	0	0	2
0	0	1	0	3
0	0	3	1	4

Table 4.2 Price matrix task dependencies

4	3	2	1	Data Center#
4	2	1	0	1
1	3	0	1	2
2	0	3	2	3
0	2	1	4	4

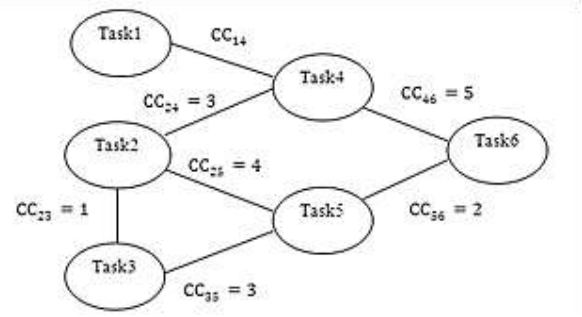
The cost function is defined as follows:

Minimize Total costs= f(x)= TM + TD + TO +TE +TC

$$TC=\sum \; \boldsymbol{Cc\,ji + Dckl}$$

Where $1 \leq j.i \leq T$, $1 \leq k, l \leq R$

	First machine	Second machine	Third machine	Fourth machine
Work	3	4	1	2



Cc, Dc are matrix elements tasks and dependent costs.

In this part, we optimize the cost function by using pso algorithm and the way of performance of each task for each car is determined to have the lowest cost. To convert this issue to a format which PSO could run it, we shoul define the limitation exactly to get correct results, So, in the next step, at least each device have a work, not in such way that many work are done by a device and another device does not have any work. For this part, we specify constraint for costs that is acceptable answer which every works will be divided equally between all equipment and correct answers should be also determined in the cost function.

The objective function described in this section. The calculation of the objective function is described below. Imagine we want to shift the task of machine number 2 to machine number 3, rows and columns of matrix hould be moved in a way which Each task is assigned to the correct machine. In this particular case, row 2 to row 3 is replaced and column 2 and the third column are obtained at the end of a diagonal matrix in this case which confirms the correctness of operations.

After this step, we calculate the cost for each machine which is the product of two tasks and dependent costs matrixes. We have used PSO for aforementioned example, The results after 20 times as follows which final composition shows

that the work assignement to the machine according to the cost function is as follows:

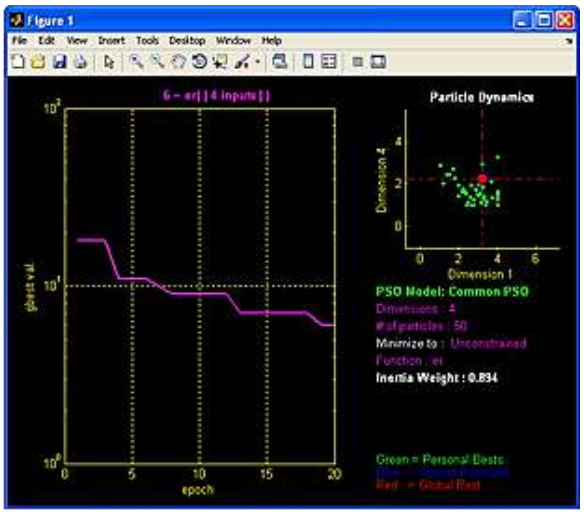


Figure 2 PSO output for the first case

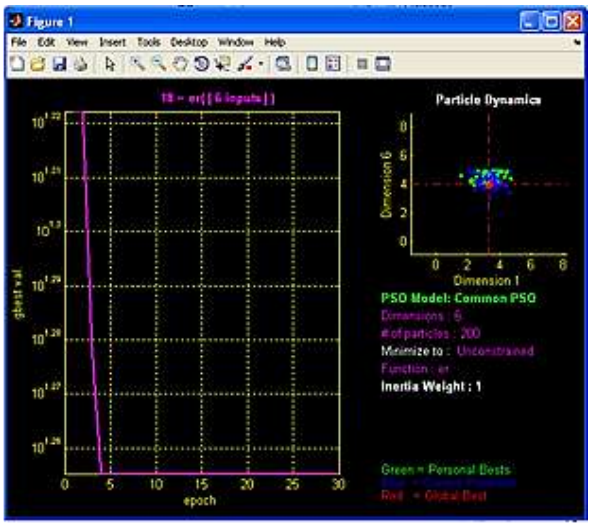


Figure 3 PSO output for the second case

	First machine	Second machine	Third machine	Fourth machine	Fifth machine	Sixth machine
Work	3	4	1	2	2	4

There is a condition in this case that must be met for the following:

Each machine can accept a maximum of two works

All machines must have partners to carry out the work.

The objective function can be expressed as before, except that dependent costs matrix is 4 x 4, When it would be expressed in terms of works, it needs to become a 6 x 6 matrix. For this conversion, It is necessary to repeat a column of the matrix twice and Or in other words, the rows and columns of the matrix which allocated to a machine that can do two things at the same time , is repeated twice and The general matrix is obtained. In this case, the simulation was conducted and the following results were obtained. Pso which have been used exerted initial population of 100 particles and we determined the number of repetitions on PSO, the amount of 30 was considered for repetitions.

4.Conclusion

According to the simulation results and adapting to the reality of tasks division can be concluded that PSO can properly diagnose the tasks division in cloud computing according to cost functions and graphs and for the division of tasks provides a good answer.

5. Refrence

[1] S.Pandey,L.Wu,S.M.Guru,andR.Buyya, A Particle Swarm Optimization-based Heuristic for Scheduling Workflow Applications in Cloud Computing Environments , peoc. On th IEEE International Conference on Advanced Information Networking and Applications, 2010 .

[2] Junjie Peng; Zhou Lei; Wu Zhang, An Energy-efficient Scheduling Approach Based on Private Clouds, Journal of Information & Computational Science 8: 4 (2011).

[3] M. Damshenas, et al., "Forensics investigation challenges in cloud computing environments," in Cyber Security, Cyber Warfare and Digital Forensic (CyberSec), 2012 International Conference on, 2012, pp. 190-194.

[4] Y. Song, H. Wang, Y. Li, B. Feng, Y. Sun," Multi-Tiered On-Demand resource scheduling for VM-Based data center", in: Proceedings of the 9th IEEE/ACM International Symposium on Cluster Computing and the Grid , Shanghai, China,pp. 148–155 Object Management Group,2009

[5] M. Armbrust, A. Fox, R. Griffith, A.D. Joseph, R. Katz, A. Konwinski,"A view of cloud computing, Commun". ACM New York, USA,pp 50–58,2009.

[6] G. Dhiman, K. Mihic, T. Rosing, "A system for online power prediction in virtualized environments using gaussian mixture models", in: Proceedings of the 47th ACM/IEEE Design Automation Conference,Anaheim, USA, pp. 807–812,2010.

[7] Ramgovind S, Eloff MM, Smith E, 2010, "The Management Of Security in Cloud Computing", IEEE.

[8] C. Clark, K. Fraser, S. Hand, J.G. Hansen, E. Jul, C. Limpach, I. Pratt and A. Warfield, "Live Migration of Virtual Machines", In Proc. of the second USENIX Symposium on Networked.

[9] David Burford, CFA, MBA, MCP Cloud Computing: A Brief Introduction February 20, 2010 .

[10] Gong, Ch., Liu, J., Zhang, Q., Chen, H., and Gong, Zh. 2010. The Characteristics of Cloud Computing. 39th International Conference on Parallel Processing Workshops.

[11] Saynsky, Barry (1390), compete reference of cloud computing, Translated by Navid Farokhi, Computer Sciences Publication, First Version, Fall 1390

[12] T. Wei-Tek, et al., "Service-Oriented Cloud Computing Architecture," in Information Technology: New Generations (ITNG), 2010 Seventh International Conference on, 2010, pp. 684-689.

[13] Rajkumar Buyya , Chee Shin Yeo , Srikumar Venugopal,James Broberg, and Ivona Brandic."Cloud Computing and Emerging IT Platforms:Vision , Hype , and Reality for Delivering Computing as the 5 th Utility".Future Generation Computer Systems, , Amesterdam,The Netherlands,pp 599-616, June 2009.

[14] M. Armbrust, A. Fox, R. Griffith, A.D. Joseph, R. Katz, A. Konwinski,"A view of cloud computing, Commun". ACM New York, USA,pp 50–58,2009.

[15] Ghazizadeh, "Cloud Computing Benefits and Architecture in E-Learning," in Wireless, Mobile and Ubiquitous Technology in Education (WMUTE), 2012 IEEE Seventh International Conference on, 2012, pp. 199-201.

[16] Nurmi, D., Wolski, R., Grzegorczyk, C., Obertelli, G., Soman, S., Youseff, L., and Zagorodnov, D. 2008. Eucalyptus opensource cloudcomputing system. In CCA08: Cloud Computing and Its Applications.

[17] K. R. Joshi, et al., "Dependability in the cloud: Challenges and opportunities," in Dependable Systems & Networks, 2009. DSN '09. IEEE/IFIP International Conference on, 2009, pp. 103-104.

[18] Buyya, R., Yeo, C.S., Venugopal, S., Broberg, J., and Brandic, I., 2009. Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the 5th utility. Future Generation Computer Systems, (2009), 25 (6), pp. 599-616.

[19]C. Clark, K. Fraser, S. Hand, J.G. Hansen, E. Jul, C. Limpach, I. Pratt and A. Warfield, "Live Migration of Virtual Machines", In Proc. of the second USENIX Symposium on Networked Systems Design and Implementation (NSDI), Boston, MA, USA, May 2005.

[20]Chunming Rong, Son T. Nguyen, Martin Gilje Jaatun, 2012, "Beyond Lighting: Asurvey on security challenges in cloud computing", Elsevier.

[21] D. Zissis and D. Lekkas, "Securing e-Government and e-Voting with an open cloud computing architecture," Government Information Quarterly, vol. 28, pp. 239-251, 2011.

[22] David Burford, CFA, MBA, MCP Cloud Computing: A Brief Introduction February 20, 2010 .

[23] Amrhein, D., Forget Defining Cloud Computing.Available[Online]:http://ibm.ulitzer.com/node/10188.

[24] Ramgovind S, Eloff MM, Smith E, 2010, "The Management Of Security in Cloud Computing", IEEE.

[25] M. T. B. Aun ,et al., "Cloud enabled spam filtering services: Challenges and opportunities," in Sustainable Utilization and Development in Engineering and Technology (STUDENT), 2011 IEEE Conference on, 2011, pp. 63-68.

[26] Gong, Ch., Liu, J., Zhang, Q., Chen, H., and Gong, Zh. 2010. The Characteristics of Cloud Computing. 39th International Conference on Parallel Processing Workshops

[27] P. Mell and T. Grance. (2009, Oct.). The NIST definition of cloud com-puting, NIST special publication 800-145, National Institute of Standards and Technology. Available: http://www.csrc.nist.gov/publications/nistpubs/800-145/SP800-145.pdf.

[28] S. Chungman, et al., "Implementation of Cloud Computing Environment for Discrete Event System Simulation using Service Oriented Architecture," in Embedded and Ubiquitous Computing (EUC), 2010 IEEE/IFIP 8th International Conference on, 2010, pp. 359-362.

[29] Ivanka Menken , Gerard Blokdijk."Cloud Computing Virtualization Specialist Complete Certification Kit".2008.

[30] S. Marston, et al., "Cloud computing — The business perspective," Decision Support Systems, vol. 51, pp. 176-189, 2011

International Journal of Computer Applications Technology and Research
Volume 5– Issue 6, 376 - 379, 2016, ISSN:- 2319–8656

A Review of Image Processing Software Techniques for Early Detection of Plant Drought Stress

Chege A. Kirongo
Computer Science and IT Department
Meru University of Science and Technology
Meru, Kenya

Abstract: Water stress is one of the most important growth-limiting factors in crop production around the world, water in plants is required to permit vital processes such as nutrient uptake, photosynthesis, and respiration. Drought stress in plants causes major production losses in the agricultural industry worldwide. There is no sensor commercially available for real-time assessment of health conditions in beans. Currently, there are several methods to evaluate the effect of water stress on plants and commonly practiced method over the years for stress detection is to use information provided by remote sensing. Studies exist which determined the effect of water stress in plants grown under the different watering regime, while other studies explore the performance of the artificial neural network techniques to estimate plant yield using spectral vegetation indices. This review recognizes the need for developing a rapid cost-effective, and reliable health monitoring sensor that would facilitate advancements in agriculture

Keywords: Image Processing, Artificial Neural Networks, Drought stress, Algorithm, Technique

1. INTRODUCTION

Water stress is one of the most important growth-limiting factors in crop production around the world. Water in plants is required to permit vital processes such as nutrient uptake, photosynthesis, and respiration. Drought stress in plants causes major production losses in agriculture industry worldwide. Monitoring of health and detection of the draught in plants and trees is critical for sustainable agriculture.

To the best of my knowledge, there is no sensor commercially available for real-time assessment of health conditions in beans. Currently, there are several methods to evaluate the effect of water stress on plants. A promising and commonly practiced method over the years for stress detection is to use information provided by remote sensing. The adaptation of remote sensing and other non-destructive techniques could allow for early and spatial stress detection in vegetables.

Studies exist which determined the effect of water stress in plants grown under the different watering regime. Other studies explore the performance of the artificial neural network techniques to estimate plant yield using spectral vegetation indices. It describes the currently used technologies that can be used for developing a ground – based sensor system to assist in monitoring health and draught in plants under field conditions.

These technologies include normalization difference vegetation index, leaf temperature and microclimatic information, automatic segmentation for scanned images to visualize features by specific color hues and intensities. The method involves classification of pixels, determined by leaf regions with or without color. Artificial Neural Network is applied in a self – organizing Kohonen network and a linear perceptron output layer. Other researchers have used ANOVA analysis of the timing of draught stress detection in plants.

2. BACKGROUND

In order to assess the vitality of a plant, it is necessary to capture the required features which describe the plant's state. For instance, the bending angle of the leaves towards the ground might be a measure of the drought stress level of the plant. The reason is the loss of turgor pressure caused by the decline of water in the leaves. This effect forces the leaves to bend down, due to gravity. Features like this can be used in biological models in order to describe how certain feature values correspond to the health state of a plant. Unfortunately, relevant features are often difficult to measure using remote systems. There are two main difficulties:

There are two main difficulties:

Firstly, the measurement system must provide sufficient data for the extraction of the features. For example, a single 2D image of a bean plant taken by a simple digital camera would not be able to obtain information about the bending angles since most leaves would be in a position where it is impossible to capture the required information. Important key points like the tip or the stem of a leaf might not be visible in all cases due to an unfavorable point of view or overlapping leaves. Taking several pictures from different points of view might solve this problem but raises others, e.g. how to identify same leaves in different images.

Secondly, the features must be extracted from the measurement data. Thus, potent imaging algorithms are required which are able to detect relevant parts of the plant and infer complex features from the measurement data. However, this is not a trivial task since the data is often very complex. In the case of laser scanners, the measurement data might consist of point clouds with thousands of 3D points. Thus, the measurement data can be very large and full of redundancies.

Nevertheless, some relevant information might be missing, e.g. because of occlusions or other negative effects.

3. IMAGE PROCESSING TECHNIQUES

3.1 Existing Techniques

There exist several methods for plant measurement. In [1], a short overview of several methods to acquire structural data is given: There are contact measurement systems, for instance, magnetic digitizers, which consist of a signal receiver and pointer, allowing the user to record the 3D spatial coordinates of the pointer within a [certain] hemisphere [...] from the receiver" [1, p. 17]. Nevertheless, magnetic digitizers are

prone to disturbances caused by iron, which might be a problem in greenhouse environments.

To overcome this problem, sonic digitizers can be used. However, those are sensitive to wind fluctuations and are therefore not suited for applications in the field [1, p. 17].

Besides this contact measurement systems, there is also a big range of non-contact measurement systems. Apart from a sheet of light laser scanning, which was used for this thesis and is therefore explained further in section 3.1, there exist systems based on time of light, volumetric intersection, stereo vision and other methods [1, p. 17_].

Time of light systems works by emitting a laser beam which is backscattered by the plant. The returning beams are then captured via a photodiode in the receiver optics. The distance between the plant and the optics can then be estimated by measuring the time interval between emitting and receiving the laser beam. In volumetric intersection systems, the surface of a plant is captured by recording the silhouette of the plant against a monochromatic background, which is then discarded. By turning the plant and taking silhouettes from different angles, the 3D surface of the plant's outline can be reconstructed. Obviously, this method has problems with occlusions or overlapping plant parts.

Stereo vision systems use two cameras, positioned at a distance, which record the same scene. From the shift of corresponding points in both views, the 3D position of these points in real-world coordinates can be computed.

All of the mentioned systems are used to capture the surface and therefore the architecture of a plant. To acquire data about functional aspects of plant growth, for instance, water transport or pigment content, other imaging techniques are used.

Chaerle [2] gives an overview of such imaging techniques which are used to monitor plant health. Those include fluorescence, luminescence, thermal imaging, magnetic resonance and reactance measuring. These imaging techniques are already proven to be useful for vitality assessment.

Please use a 9-point Times Roman font, or other Roman font with serifs, as close as possible in appearance to Times Roman in which these guidelines have been set. The goal is to have a 9-point text, as you see here. Please use sans-serif or non-proportional fonts only for special purposes, such as distinguishing source code text. If Times Roman is not available, try the font named Computer Modern Roman. On a Macintosh, use the font named Times. Right margins should be justified, not ragged.

3.2 Plant Stress Detection Techniques

As it was shown in the previous section, there exists a wide range of different imaging systems, each being useful for different data acquisition tasks. However, further techniques are necessary to detect valuable information in the data and extract it reliably. Depending on the task, each has a varying degree of difficulty, which is illustrated in the following examples.

In the experiments described by Romer et al. [3], hyperspectral images of barley and corn plants have been analyzed in order to detect drought stress in a very early stage, i.e. before the plant is damaged. The experiments have been performed on the level of single plants in the case of barley, whereas corn plants were measured directly in the field.

The data consisted basically of histograms of the hyperspectral plant images. To assess the vitality state of the plants, Romer et al. used an unsupervised classification method called simplex volume maximization. The key idea of this method is to find archetypal samples in the data which represent the most extreme cases of plant growth, e.g. very healthy or very stressed plant samples.

This is achieved by comparing the distances between all data points and declaring the samples as archetypes which span the greatest volume of the distance around all other samples. Afterwards, all non-archetypal samples can be expressed as a convex combination of the archetypes, where the coefficients relate to the similarity of a data sample to a respective archetype. The benefit of this is that the resulting classes can be interpreted by humans although they were learned using an unsupervised learning technique. Nevertheless, at least with respect to the barley dataset, a human expert was necessary in order to assert the correctness of the archetypes.

The method was compared to visual classification and to the usage of vegetation indices, e.g. calculating the ratio of ground which is covered by biomass. Romer et al. achieved good results. In the barley experiment, stress was detected 5 days earlier than with visual classification and 4 days faster than with the tested vegetation indices.

As Romer et al. [3, p. 879] mention, despite several laboratory studies that have shown a relationship between the amount of water in the leaf and the spectral reactance in the optical region [4], at canopy level the determination of water content presents some difficulties, mainly due to the large reactance variation among leaves with the same water status [5], structural changes associated with loss of turgor [6] or small reactance differences at different levels of water stress".

Therefore, it is difficult to detect drought stress in single plants using reactance imaging.

Moreover, since the category of plants is grown to be used for medical products, it is necessary to control the state of every single plant regularly and very often, e.g. every day. In the corn experiment of Romer et al., however, the measurement system took one minute to acquire an image of the size _2x1.5m, which is quite slow if a large field with hundreds of plants has to be processed.

4. DISCUSSION

4.1 State of the Art in Plant Drought Stress Algorithms

A different kind of approach is taken by Seatovic [10], who presents a system for the recognition and treatment of broad-leaved dock, which is a highly competitive and persistent sort of weed. This system uses an infrared laser scanner and a high-resolution smart camera which are mounted on a carrier vehicle. Besides the sensor system the vehicle carries an herbicide spraying component which is able to treat certain parts of the captured area, if a weed plant is detected. The vehicle can be moved over a field with a speed of 1ms 1 while weed plants are detected and treated in real-time. In the image data, weed plants are identified by finding contiguous surface patches. This is done with a simple edge detection algorithm [10, p. 174f]. The surface patches are then compared to objects of a plant database with respect to simple features like shape and texture. The main benefit of Seatovic's approach is that features like shape can be derived from the raw measurement data. Therefore, they can be computed fast

enough to be used in a real-time system. Nevertheless, he reports that only 62-91% of the weed plant surface have been detected in field tests. This detection rate was sufficient in his project since the radius of the herbicide spraying nozzles was big enough to treat all weed plants even if not all leaves have been detected.

However, other application contexts might require a higher accuracy. For example, a system for treatment of vermin might need higher accuracy rates since any missed bugs could spread across already treated plants.

Both approaches report good results on the level of whole plant stands, whereas a reliable detection of single leaves is difficult. A reason might be that both approaches use quite simple features. Although they differ in terms of data acquisition and feature extraction methods, both approaches use simple features which rather describe the images instead of the plants which are depicted.

Consequently, it is difficult to derive information about certain plant parts like leaves since they are on a higher level of detail. If that information is required, further advanced feature extraction techniques are necessary. Unfortunately, the segmentation of image data to corresponding plant parts and the automatic calculation of certain features is a challenging task due to its complexity. As van der Heijden [11, p. 19, p. 22] stated, "automatic feature extraction is still in its infancy" and consequently "plant features still have to be extracted interactively since current software is not yet capable of fully automatic extraction of plant features in complex images".

Since an advanced feature extraction is the bottleneck of many applications, there are some recent approaches to a solution. Paulus et al. [12] provide an algorithm to classify parts of point cloud data, e.g. to distinguish leaves from stems. Their approach is based on surface feature histograms, which consist of local geometric point features, e.g. neighborhood characteristics. By differentiating among different classes of histograms, different plant parts can be detected in the point cloud. They confirmed their methods by successfully applying them on grapevine and wheat plant organs, where classification accuracy rates up to 98% could be reached.

In the work of Balfer [13], a skeletonization algorithm is explained, which uses semantic annotations to extract a precise 3D model of complex grapevine stem systems. From this model, structural features can be extracted, e.g. the lengths of the peduncles. However, the approach requires expert knowledge in order to find and model suitable semantic annotations

4.2 Existing Methodological Approaches

Dornbusch et al. [14] propose a method to extract morphological traits of a plant, e.g. width and length of stem segments and leaves by fitting an architectural model to point cloud data. This method is quite promising since it provides meaningful features which can be interpreted by human experts. However, a crucial step, namely segmenting the point cloud of a whole plant into smaller segments (i.e. the actual detection), is still performed manually in Dornbusch's approach.

Camargo [15] uses histogram-based approaches to segmenting 2D images of cotton plants, e.g. to find spots which were damaged by bugs or infected with diseases. The approach works well and is feasible, nevertheless, it is only applicable to 2D image data and is therefore not suitable to extract 3D features like bending angles of leaves.

Given that meaningful features of plant growth can be extracted from sensor data, the actual assessment of plant vitality can be performed by assigning certain classes or measures of plant health to patterns or constellations of attribute values. In many agricultural applications, standard machine learning techniques are used to deal with this kind of task.

Mucherino [16] provides a survey of how techniques such as k-nearest neighbor, artificial neural networks or support vector machines have been successfully applied to solve agricultural problems. The fields of interest include prediction of wine quality, soil quality analysis, recognition of pig sounds and the detection of meat and bone meal in feedstuffs for animals. Unfortunately, stress classification of plants is none of the discussed problems.

In the field of plant classification, support vector machines are widely used. Rumpf et al. [18] use support vector machines to distinguish diseased from non-diseased hyperspectral sensor data of sugar beet leaves. They mainly use simple features like vegetation indices, i.e. the ratio of ground which is covered by biomass. However, since it is based on hyperspectral data, their approach allows for early disease detection.

Camargo [9] uses support vector machines to classify visual symptoms of cotton plant diseases, i.e. visual damages of the leaves. As it was already mentioned, his approach relies on 2D images which are searched for relevant segments using color intensity histograms. Therefore, the approach is only feasible if no other information, e.g. 3D geometry, is required.

4.3 Approaches for Morphological Design of Leaf Model

Most of the mentioned approaches tend to avoid the bottleneck of feature extraction by reducing it to a minimum, either by using global image data like spectral histograms or by only extracting simple features like texture, color or width and length values of simple shapes. However, those simple features only describe visual properties of the images but lack a semantic relation to the growth behavior of the plants which are depicted.

More desirably, the features should correspond directly to information which is interesting for a human expert. Therefore approaches that links to the work of Uhrmann et al. [5], in which a morphological leaf model should be studied to extract feasible growth parameters. The difference to other approaches is that the model should consider complex features, e.g. attributes describing the bending behavior of the leaf along certain axes. In contrast to simple attributes like length or width of the leaf, this complex attributes should be directly correlated with the vitality state of the plant.

5. CONCLUSION

Since using a leaf model is a very new technique, a model should be validated so as to analyse the leaf model features and to show which of them are relevant for the description of drought stress with respect to plants.

Furthermore, classification techniques should be examined for the evaluation of the model leaf features. Many of the mentioned approaches use support vector machines for the classification process since they provide excellent results in most cases. However, there are different classification techniques which are easier to interpret by human experts. For example, decision trees or linear regression provide an easier model in terms of comprehensiveness.

Therefore, research to evaluate several classifiers with respect to the task of learning the features provided by the leaf model can be considered. For instance, if decision trees turn out to provide classification results comparable to support vector machines, they should be preferred in a practical application, since they are easier to evaluate by non-technical experts, e.g. biologists.

Therefore, a model for the impact of drought stress on plant growth can be inferred from measured geometric leaf features using machine learning techniques. The features were acquired using a measuring system developed at the Fraunhofer IIS. The aim was to evaluate which of the available features are required to model drought stress and whether they are salient to reach a reliable classification performance.

An overview of existing approaches and techniques with respect to vitality assessment of plants via remote sensors should be discussed.

6. ACKNOWLEDGMENTS

Thanks to the Computer Science and Information Technology graduate staff and students.

7. REFERENCES

[1] van der Heijden, G., de Visser, P., and Heuvelink, E. (2007). Functional-Structural Plant Modelling in Crop Production: Measurements for Functional-Structural Crop Models, chapter 2, pages 13{25. Springer.

[2] Chaerle, L. and Van Der Straeten, D. (2001). Seeing is believing: imaging techniques to monitor plant health. Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression, 1519(3):153166.

[3] Romer, C., Wahabzada, M., Ballvora, A., Pinto, F., Rossini, M., Panigada, C., Behmann, J., Leon, J., Thurau, C., Bauckhage, C., Kersting, K., Rascher, U., and Plumer, L. (2012). Early drought stress detection in cereals: simplex volume maximisation for hyperspectral image analysis. Functional Plant Biology, 39:878{890.

[4] Russell, S. J. and Norvig, P. (2010). Artificial Intelligence: A Modern Approach. Pearson Education, 3rd edition.

[5] Uhrmann, F., Hugel, C., Paris, S., Scholz, O., Zollhofer, M., and Greiner, G. (2013). A model-based approach to extract leaf features from 3d scans. In Proceedings of the 7th International Conference on Functional-Structural Plant Models, Saariselka, Finland.

[6] Mitchell, T. (1997). Machine Learning. McGraw-Hill International Editions.

[7] Han, J. and Kamber, M. (2005). Data mining: concepts and techniques. Kaufmann, San Francisco [u.a.].

[8] Alpaydin, E. (2004). Introduction to Machine Learning. Adaptive computation and machine learning. MIT Press.

[9] Camargo, A. and Smith, J. (2009a). Image pattern classification for the identification of disease-causing agents in plants. Computers and Electronics in Agriculture, 66:121{125.

[10] Seatovic, D. (2008). 3d-object recognition, localization, and treatment of rumex obtusifoliol in its natural environment. In 1st International Conference on Machine Control & Guidance.

[11] van der Heijden, G., de Visser, P., and Heuvelink, E. (2007). Functional-Structural Plant Modelling in Crop Production: Measurements for Functional-Structural Crop Models, chapter 2, pages 13{25. Springer.

[12] Paulus, S., Dupuis, J., Mahlein, A.-K., and Kuhlmann, H. (2013). Surface feature based classification of plant organs from 3d laser scanned point clouds for plant phenotyping. BMC Bioinformatics, 14:238.

[13] Balfer, J., Scholer, F., and Steinhage, V. (2013). Semantic skeletonization for structural plant analysis. In Proceedings of the 7th International Conference on Functional-Structural Plant Models, Saariselka, Finland.

[14] Dornbusch, T., Wernecke, P., and Diepenbrock, W. (2007). A method to extract morphological traits of plant organs from 3d point clouds as a database for an architectural plant model. Ecological Modelling, 200:119{129.

[15] Camargo, A. and Smith, J. (2009b). An image-processing based algorithm to automatically identify plant disease visual symptoms. Biosystems Engineering, 102, Issue 1:921.

[16] Mucherino, A., Papajorgji, P., and Pardalos, P. M. (2009). A survey of data mining techniques applied to agriculture. Operational Research, 9, Issue 2:121{140.

[17] Rumpf, T., Mahlein, A.-K., Steiner, U., Oerke, E.-C., Dehne, H.-W., and Plumer, L. (2010). Early detection and classification of plant diseases with support vector machines based on hyperspectral reflectance. Computers and Electronics in Agriculture, 74:91{99.

International Journal of Computer Applications Technology and Research
Volume *– No.*, _____ 2012

Providing a network encryption approach to reduce end-to-end delay in MANET

Simin Moghiseh,
Department of Computer Software
Ayatollah Amoli Branch, Islamic Azad University, Amol, Iran

Seyed Javad Mirabedini,
Department of Computer Software
Central Tehran Branch, Islamic Azad University, Tehran, Iran

Abstract:

A lot of research has been done on different coding techniques and benefits of their use in wired networks Since network coding was raised as a basic method for increasing network outpouring and reaching the capacity of networks. However, wireless networks are suffering from low operating power as the use of NC concept in MANET principally improves throughput rate in the wireless network .

The delay can be considered as an important parameter in networks and system delayed is not acceptable in these networks. However, the acceptable delay depends on the application although the efficiency and throughput leads to an increase in network coding , a reduction in bandwidth consumption, and a delay in sending packets is reduced by using network coding. In this study, a method is proposed for coding in the MANET, decreasing the number of sent packets, leading to a reduction in that delay .

Keywords: network coding, wireless ad hoc networks, simulation, directional antennas, simulation

1. INTRODUCTION

Mobile ad hoc wireless networks or more simply, Mobile ad hoc networks (MANET) is a set of nodes scattered geographically, which is linked with each other through a wireless medium [1]. A mobile ad hoc network does not have any wiring framework and communications are limited by battery power.

A mobile ad hoc network is a collection of mobile Telecommunications hosts, created without using a network infrastructure, and a temporary network is established by Such ad-hoc networks can be used in remote areas or areas with difficult conditions where there is no infrastructure. Furthermore, small ad-hoc networks are useful and affordable for stations in which the use of existing infrastructure is costly or provides a weaker performance than direct communication.[2].Network coding is a method used to enhance the communication networks power. Network coding is derived from the idea of hybrid packets and sending them to achieve higher throughput. Some of these methods send artificial delay to information packets [3]. The idea of network coding is derived from a clever combination of packets, from different altitude origins), and sending them in order to achieve higher throughput. At the earlier studies done on the area of network coding, improving wired networks was initially considered, although the concept of network coding could later attract the attention of researchers to improve throughput rates and optimum use of resources of wired networks. Before network coding, the intermediate nodes such as routers and switches was only responsible for storing, routing and sending the packets to the destination; therefore, in network coding, encoders are used instead of routers and switches, allowing us encode the packets and send them for intermediate nodes.

In Section two, the related works are reviewed andthe algorithm is proposed in section 3. This chapter also describes how to calculate sending packets to neighbor nodes and navigator node, as well as the coding messages received by the node navigator, and sending a message to all the nodes. Experimental results of simulation are compared with previous methods in section 4 . Finally, section 5 is related to the summary and conclusion about the material presented, and suggestions for further research.

2. Related works

In this section, the related works are briefly described.

SPIN method

The SPIN method is a family of adaptive protocols, which is able to scatter data among sensors in a sensor network effectively with the limited energy resources. The nodes run by SPIN communication protocol call their data by using a quasi-data which are descriptors with a high level.[4] Further,, in this method, nodes are using quasi-data negotiation to remove additional data on the network. In addition, nodes SPIN can decide to conduct their communications based on the information on the application and the information on the available resources . This makes the sensors be able to disperse the data efficiently in spite of their limited resources.

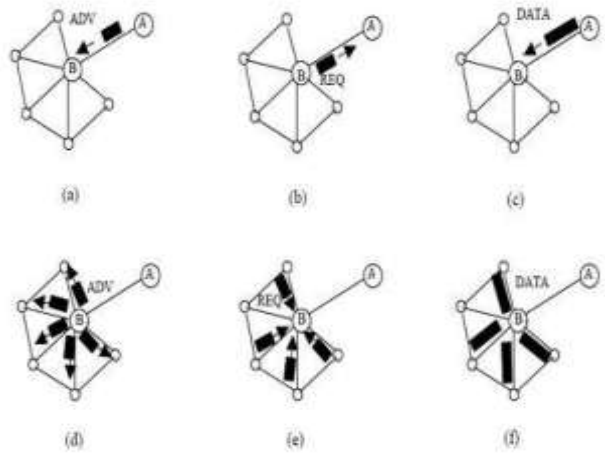


Figure 1: SPIN messages

Direct diffusion method

This protocol is considered as one of the most crucial of data-based protocols, which many protocols have been established based on . The protocol is designed so that each time a new application could be created, and the routing could start based on this new application. Each sensor receiving the request, keeps it in its memory for later use. Sensors combine data locally, leading to a reduction in the amount of transmitted information . Further, each sensor receiving a request, transfers it to its neighbors, so that gradient could be formed among them.[5]

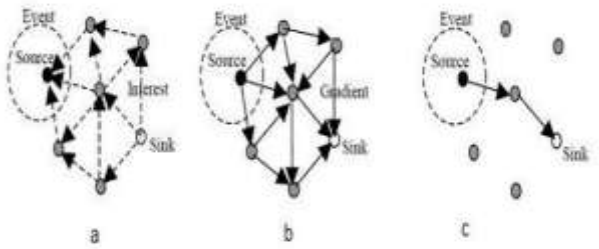


Figure 2: Direct diffusion method

Opportunistic summarizing

In general, numerous methods have been presented for summarizing the information. Among these methodsthe method of random summarizing was proposed based on the diffusion method. Based on this method, when the data are collected by sources and sent toward the recipients, intermediate nodes store information related to each other, by using filters at the application level. Then, the intermediate nodes eliminate duplicate information, or create a slight delay in sending information, and summarize the information collected from various sources. In this way, a tree of resources is formed toward a recipient as tree root. if the same data are collide with each other during the integration of the tree branches, only one of them will be sent to the root of the tree.

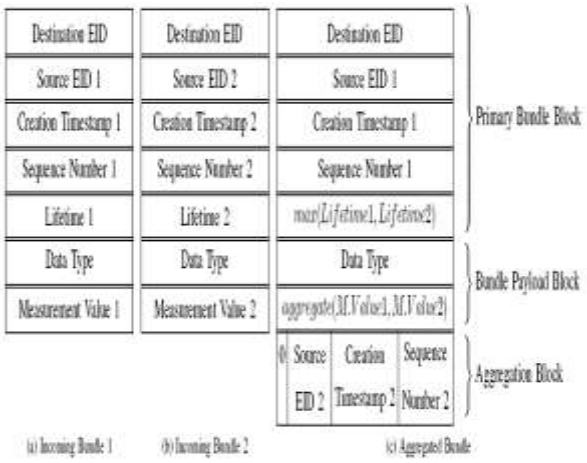


Figure 3: Opportunistic summarizing

Cope Method

Network encryption was the first architecture , which supported unicast traffic encryption, from different altitude origins, before sending them on a wireless network. Network programming as a modern technology can improve the performance of network communications. In such a network programming, hybrid packets are related to the network performance. Accordingly, several methods were proposed for the purpose of increasing throughput, based on the network programming. Many of these techniques of delay in sending the packet are used to increase throughput. However, this method is not appropriate for the applications sensitive to delay. In order to solve this problem, some methods were proposed for tradeoff between delay and throughput. A packet which is able to tolerate delays more than others has a higher priority for programming based on end-to-end delay, and programming with higher priority. Each node places packets from different flows in the virtual queue based on its next node.[6,7]

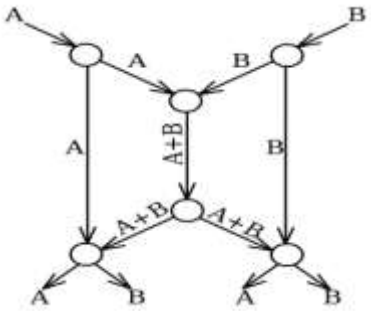


Figure 4: COPE coding

EBCD Method

EBCD algorithm takes the advantage of directional antennas for coding, and tries to reduce the number of packets sent, and for this purpose it stores the packets until the timer is expired, and is coded base on the order of their arrival.

Proposed algorithm

In this study, a new algorithm for coding is presented, aiming to reduce the number of packets sent and reduce delay. The algorithm is based on this idea that, while coding, each node can determine its neighbors, and when the packets are received uses the table for this kind of activity. Network coding allows the navigator nodes to merge incoming data, and then send it, through which the total number of packets sent is reduced. In this new design, each node sends its message to its adjacent nodes and navigator node.[8]

How does the proposed algorithm work?

The proposed algorithm works as follows:

Using the following equation, the lowest-cost distance between nodes with navigator node is selected, in order to select the set of nodes covered by the navigator node.

$$\min of \ \sum_{(i,j)\in A} a_{(i,j)} z_{(i,j)}$$

Input of this algorithm is the status of the group / message table, which is sorted in an ascending order.[9]

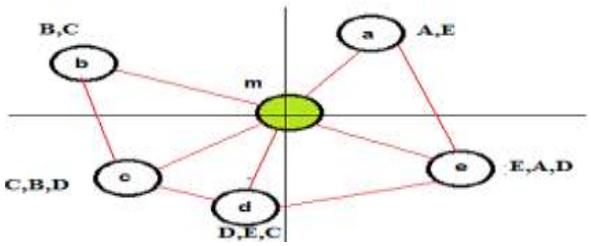


Figure 5: The network sub-graph with domain

Table 1: Table of node / Messages sorted

	A	B	C	D	E
a	1	0	0	0	1
b	0	1	1	0	0
c	0	1	1	1	0
d	0	0	1	1	1
e	1	0	0	1	1

- This table is sorted by the number of incoming packets nodes.
- To make a coded packet, the node which has received the lowest packet is prioritized.
- A not-received packet is selected from the selected node and is placed in the P set, followed by the next node.
- In the new node, an unknown packet is selected and is added p in accordance with the following conditions.
- If the selected packet already exists in P, a packet is not selected from the above node, and the next node is followed.
- If the selected packet did not exist in P, it is added to P,if all nodes can be decoded after addition ; in such cases, it is added to P, otherwise another packet in that node is taken.
- Each node can only decode those packets having at least one of the packets beforehand.
- The above steps are repeated until p sets are aerated for coding; and all packets are in the coding set.
- After coding packets, the navigator node checks to see which node should be done for sending.
- After receiving the packet, decoding operations are per[illegible]ed by nodes, and each node sends a messag 1 ts neighbors.

```
Sorting matrix;
While (zero in matrix)
Begin
        for i=1 to n /* I row number
        begin
                A= select zero in matrix;
                for j=I to n
                begin
                        B=  select zero in matrix;
                        coding=A+B;
                        if all nodes can receive (A+B)
then
                        begin
                                A+B is current coding;
                                Break;
                        End
                end
                Update matrix(coding)/* zero convert
one for all coding item
        End
end
```

Simulation

MATLAB software is used for simulation, by which a network with up to 100 nodes was considered, and the proposed method was compared with S-EBCD. For comparison, the items of packets sent, sending and delay time is used.

Simulation Scenario

In the simulation, each node has a packet to send, and the packet must be received by other nodes. In each sub-graph, there is a navigator node, which is responsible to send packets. In the proposed method, each node sends the relevant packets to adjacent and navigator nodes.[10,16]

The number of packets sent

In this case, the number of nodes up to 10 are taken into account, which increases up to 100 and the number of packet sent is computed every step, resulting in the following figure.[11]

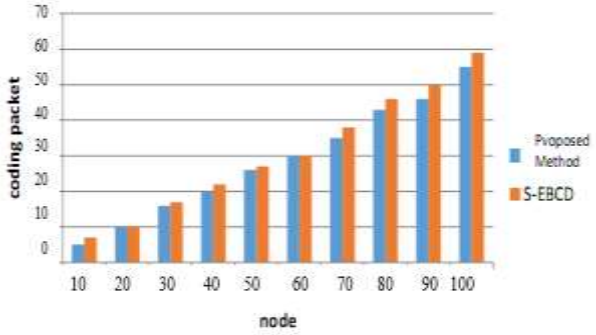


Figure 6: A Comparison of the number of packets sent

As it is evident from the simulation results in Figure 6 , in the proposed method, the number of sent packets has dropped to 11% , which can be regarded as the main reasons for the suitable coding of the Proposed method. This kind of reduction takes less time to send all packets, and there is not much delay.[13,14]

The time related to send packets

In Figure 7, the time related to send packets are compared. As it is observed from this figure , the time has reduced in the Proposed method.

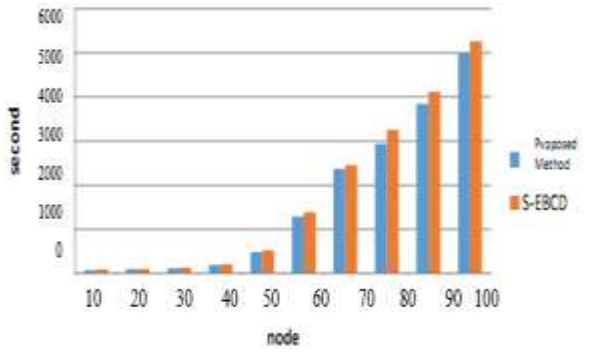


Figure 7: A Comparison of sending time

In the proposed method, the time has declinedto 13% . In Figure 8, the coding and sending time has been compared in conditions in which the number of nodes are fixed and the number of messages sent by each node are increased.[12,15]

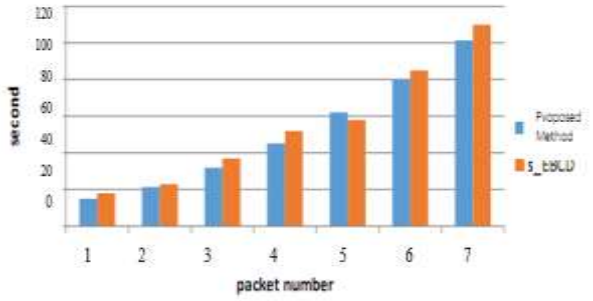


Figure 8: Comparison of sending time

In Table 2, the proposed method and S-EBCD method are compared.

Table 2- Comparison of Prposed and S-EBCD methods

	The number of packets sent	Sending time	Computational overhead	Memory used
New Method	Low	Low	Medium	Medium
S-EBCD	High	High	High	Medium

3. Conclusion

The proposed method in the current study acts in such a way that the coding packets are created in the best possible way. Based on this new method the number of packets sent are reduced, leading to a reduction in delay. The concept of network coding has attracted the attention of all the researchers, interested in improving the use of wired and wireless networks resources. Network coding networks were proposed as a solution to improve the wired networks. However, wireless networks are suffering from low operating power. Further, the concept of NC on wireless networks has improved wireless networks fundamentally throughput rate. To date, many solutions have been proposed in order to improve the performance of network coding. Finally, coding tree messages, and reducing the time and the number of packets can be considerd as a new area of research in future.

4. Refrence

Nikoee Saravani R and Mirabedini SJ , New jitter-aware network coding approach for packet selection in wireless mesh network using Jacobsen algorithm, Management Science Letters,Vol 4,pp 2487–2494,2014

Atilla, E., et al; control for intr-session network coding; IEEE Transactions onNetwork Coding, Vol 6 ; pp 719-114; 200.

Lun, D. S., et al; On coding for reliable communication over packet networks; subtted to IEEE Transactions on Information Theory;pp 2745-2759, 2005.

Bertsekas, D. P.; Network Optimization: Continuous and Discrete Models, Belmont, MA: Athena Scientific; pp 567-571; 2000.

Lun, D. S., et al; Network coding with a cost criterion, in Proc. International Symposium on Information Theory and its Applications; pp 1232–1237; 2004.

Ho, T., et al; Dynamic algorithms for multicast with intra-session network coding, in Proc. 43rd Annual Allerton Conference on Communication, Control, and Computing; pp1391-1412, 2005.

Wu, Y., et al; Information exchange in wireless networks with network coding and physical-layer broadcast, in Proc. Conference on Information Sciences and Systems ;pp 1476-1479, 2005.

Bhadra, S., et al; Min-cost selfish multicast with network coding, IEEE Trans Transactions on Information Theory, Vol 6; pp 5177–5117; 2006.

Li, Q., & Mills, D. L. (2008). Jitter-based delay-boundary prediction of wide-area networks. Networking, IEEE/ACM Transactions on, 9(5), 578-590.

Ying, L., Yang, S., & Srikant, R. (2007, April). Coding achieves the optimal delay-throughput tradeoff in mobile ad-hoc networks: Two-dimensional iid mobility model with fast mobiles. In Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks and Workshops, 2007.WiOpt 2007. 5th International Symposium on (pp. 1-10). IEEE.

Dong, Q., Wu, J., Hu, W., & Crowcroft, J. (2007, September). Practical network coding in wireless networks. In Proceedings of the 13th annual ACM international conference on Mobile computing and networking (pp. 306-309). ACM.

WiCom'09. 5th International Conference on (pp. 1-4). IEEE. Yeow, W. L., Hoang, A. T., & Tham, C. K. (2009, April). Minimizing delay for multicast-streaming in wireless networks with network coding. In INFOCOM 2009, IEEE (pp. 190-198). IEEE.

Distributed utility maximization for network coding based multicasting: a shortest path approach. Wu, Y. and Kung, S.-Y. 8, IEEE Journal on Selected Areas in Communications, Vol. 24, August 2012, pp. 1475-1488.

Dong, Z, Shen,Li,Y. Dynamic Network Coding with Delay-guarantee in Multi-hop Wireless Networks. Journal of Computational Information Systems, Vol (4), Jan 2013, pp.1529-1538.

V. Handziski, A. K¨opke, H. Karl, C. Frank, W. Drytkiewicz, “Improving the Energy Efficiency of Directed Diffusion Using Passive Clustering,” European Workshop on Wireless Sensor Networks 2012 (EWSN 2004), pp. 172–187, 2012.

John Heidemann, Fabio Silva, Yan Yu, Deborah Estrin, and Padmaparma Haldar. “Diffusion Filters as a Flexible Architecture for Event Notification in Wireless Sensor Networks.” Technical Report ISITR-556, SC/Information Sciences Institute, April, 2002

International Journal of Computer Applications Technology and Research
Volume 5– Issue 6, 385 - 390, 2016, ISSN:- 2319–8656

Simulation Based Analysis of Bee Swarm Inspired Hybrid Routing Protocol Parameters Using Realistic Mobility Model In Vehicular Ad-Hoc Networks

Kamaljeet Kaur
Department of Computer Science and Engineering
G.G.S.C.E.T, Guru Kashi University
Talwandi Sabo, India

Abstract: Vehicular Ad-hoc Networks (VANET's) are basically emanated from Mobile Ad hoc networks (MANET's) in which vehicles act as the mobile nodes, the nodes are vehicles on the road and mobility of these vehicles are very high. The main objective of VANET is to enhance the safety and amenity of road users. It provides intelligent transportation services in vehicles with the automobile equipment to communicate and co-ordinates with other vehicles in the same network that informs the driver's about the road status, unseen obstacles, internet access and other necessary travel service information's. The evaluation of vehicular ad hoc networks applications in based on the simulations. A Realistic Mobility model is a basic component for VANET simulation that ensures that conclusion drawn from simulation experiments will carry through to real deployments. This paper attempts to evaluate the performance of a Bee swarm inspired Hybrid routing protocol for vehicular ad hoc network, that protocol should be tested under a realistic condition including, representative data traffic models, and the realistic movement of the mobile nodes which are the vehicles. In VANET the simulation of Realistic mobility model has been generated using SUMO and MOVE software and network simulation has been performed using NS2 simulator, we conducted performance evaluation based on certain metric parameters such as packet delivery ratio, end-to-end delay and normalized overhead ratio.

Keywords: Vehicular Ad hoc Network (VANET), Bee swarm routing protocol, Genetic Algorithm, Realistic mobility model, SUMO, MOVE, NS-2, Simulations.

1. INTRODUCTION

Vehicular Ad hoc Network (VANET) is a new revolutionary phase of automotive industry. Today, a vehicle is not just a thermo mechanical machine with few electronic devices; rather, recent advancement in wireless communication technologies has brought a major transition of vehicles from a simple moving engine to an intelligent system carrier.[1] The components of VANET mostly resembles the MANET features like as the techniques of self-configuring, self-management and autonomous wireless network. A vehicular ad-hoc network is a technology that uses moving vehicles as nodes in a network to establish a arbitrary mobile network. VANET provides communication between adjacent vehicles and between vehicles and nearby fixed point and offer some intelligent activities. It is an intelligent network of vehicles, called Intelligent Transportation System (ITS) which is used to ensure the security services of driver assistance and comfort of road users. Intelligent Transportation System include all types of communication in vehicles, between vehicles-to-vehicles (V2V) and Vehicle-to-Infrastructure (V2I).[2] The basic target of these communications is to increase the road safety, provide the better driving surroundings and cautioning messages need to be sent to the vehicles for different functions through the inter-vehicle communications.

VANET is a subset of Mobile Ad hoc Network (MANET) which consists of number of vehicles with the ability to communicate with each other, they transmit information via radio in order to enhance the security and efficiency of transportation system, and allow internet access to the users. It is supposed that each vehicle has a wireless communication equipment to provide ad hoc connectivity. Such networks comprise of sensors and On Board Units (OBU) installed in the car as well as Road Side Unit (RSU). The data collected from sensors on the vehicles can be displayed to the drivers and sent to the RSU or even broadcasted to other vehicles depending on its nature and importance.[3]

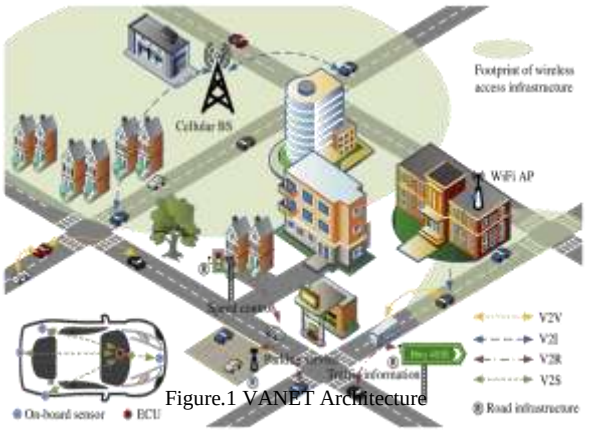


Figure.1 VANET Architecture

One of the main challenges in the full deployment of VANET is the testing of routing protocols and algorithms because vehicular ad hoc networks topologies are very different by

nature as it changes very frequently and rapidly, basically it depends on the nature of movement of vehicles, scopes and scale of VANET networks. So, effectively design routing protocol with a more intelligent use of resources is a good way to cover these problems and able to improve the performance of the transmission and dissemination of information which are made according to the routing protocols.

In this paper, we proposed a Hybrid Bee Swarm Routing Protocol which is developed by using Bee swarm routing protocol as well as Genetic Algorithm simultaneously that can efficiently handle the variations of network parameters to keep its performance constant over network changes. For testing and evaluating this proposed protocol we used Realistic Mobility Model. The high mobility of nodes in VANET makes mobility model selection one of the most important parameters when evaluating any protocol. Mobility models helps to determine the location of nodes in the topology at any given instant, which strongly affects network connectivity and throughput.[4] The performance of the proposed protocol has been studied using simulation tools SUMO and MOVE and also Network Simulator(NS2).

2. RELATED WORK

Vehicular Ad hoc Network (VANET) present various services to the road users to communicate and transmit information with each other to help out the vehicle drivers to avert any kind of critical situations with providing secure and safely environment. For the transaction of data information between two entities in a network we need to use an efficient routing protocol to manage the performances of vehicles at different network parameters. In this paper, we introduced a Bee Swarm Hybrid Routing Protocol for the improvement of the routing efficiency, to keep its performance constant whenever the network parameters changes. The proposed protocol is basically a combination of two techniques that is Bee swarm routing protocol and Genetic Algorithm. Bee Swarm routing protocol is describe the intelligent foraging nature of honey bee swarm which is consists of three groups: packers, scouts and foragers. Honey bees use distinct dancing techniques for communication which is performed by the scout bees like as waggle dancing and round dancing. If the food is closure to the hive, bees dance in circular motion that is known as round dancing and the waggle dance is the dance when bee repeated a same move in a form of number eight which implies the distance and the direction of food. This protocol is able to vigorously find out multiple paths between source and destination, and also helps to distribute traffic across them.

Other technique used in proposed system is the Genetic algorithm. Genetic Algorithm is heuristic technique that implies the methods to detect the ideal solutions which is based on the assumptions of natural evolution and genetics. In this proposed system we use these two techniques simultaneously in order to establish the hybrid algorithm. In this algorithm the individual (is set of solutions available for particular problem) refers to nodes that are participating in the route from source to destination, Population indicate the group of individuals that are selected as participant for the routing process from source to destination. Fitness function is used to identify the compatible solution for the problem; if the evaluated individuals execute their functions in appropriate way then they got chances to be selected as a fittest candidate node for routing from source to destination.

This proposed algorithm will provide the reliable and efficient routes from the source node to the destination node with better performance results. It is not possible for directly implementation of this protocol due to the cost and complexity. So, the simulation becomes the tool of choice to evaluate these solutions. In this paper, we used a Realistic Mobility Model for the simulation of the evaluated routing protocol which characterized the real world region. These Mobility model represents the flow diagram of mobile users, including its location, speed and acceleration over time.[2] In Realistic Mobility Model the proposed protocol should be tested and evaluated in realistic conditions like as the movement of the nodes which are vehicles and assure that the achieved results will carry through the real world zone.

3. SIMULATION METHODOLOGY

The Simulation Methodology is a very crucial way to formulating the realistic evaluation of the network; it provides the manifest idea of the involved research. So, choosing the appropriate methodology is very necessary as it will enhance the performance of this system. In this paper, we are going to use Simulator for Urban mobility software (SUMO) and MOVE (Mobility model generator for Vehicular networks) and also Network Simulator (NS2.35) software to run simulations. By using SUMO software, we created a Realistic Mobility Model for Vehicular Ad hoc Networks (VANET'S).

3.1 Mobility Model Generator for Vehicular Networks (MOVE)

Mobility model generator for vehicular networks (MOVE) is software to facilitate users to rapidly generate realistic mobility model for VANET simulations. MOVE is a java-based application built on SUMO with a facility of GUI. It provides an environment that allows users to quickly pinpoint incorrect details and manage details overhead. MOVE is built on the top of an open source micro-traffic simulator SUMO.[6] The output of MOVE is a mobility trace file that consists the information of realistic vehicle movements which can be instantly used by the network simulators such as NS2 and qualnet. It allows users to generate realistic mobility models by two approaches: interfacing with real world data map and other one is with Google earth.

3.2 Simulation of Urban Mobility (SUMO)

Simulation of Urban Mobility (SUMO) is an open source traffic simulation package including net import and demand modeling components. SUMO helps to investigate several topics e.g. route choice and traffic light algorithm or simulating vehicular communication. Therefore the framework is used in different projects to simulate automatic driving or traffic management strategies. [7] It allows the user to build a customized road topology, in addition to import the different readymade map formats of many cities and towns of the world. The map area which is to be simulated is exported from OpenStreetMap(OSM) which is saved as .osm file. This .osm file converted into .net.xml file. The route file and

.net.xml file then transformed into sumo trace.xml file using the configuration file of sumo, this file is changed into NS2 file using trace Exporter.py of SUMO which is used for the simulation of mobility model nodes of VANET.

3.3 Network Simulator (NS 2.35)

Network simulation is a widely used tool to simulate the behavior of wired and wireless networks. It is a technique where a program models the behavior of a network either by calculating the interaction between the different network entities. NS is a discrete event simulator targeted at networking research. Ns provides substantial support for simulation of TCP, routing and multicast protocols over wired and wireless (local and satellite) networks.[8] The scripting of NS is written in C++,with an Otcl interpreter. In our case 2.35 version of network simulator is used which is commonly called as NS2.

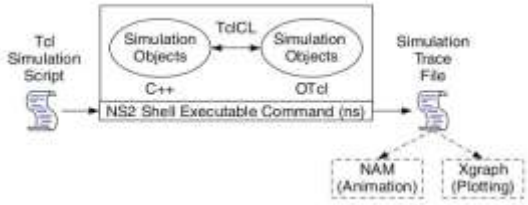


Figure.2 Basic Architecture of NS

4. SIMULATION SETUP

The VANET network is simulated using SUMO and MOVE and also NS -2.35 to analyze the performance of the proposed protocol parameters. The Simulation parameters that are used in this simulation given below in the following table:

Table I : Simulation Setup

S.No.	Parameters	Values
1	NS version	ns-allinone-2.35
2	MOVE version	2.92
3	SUMO version	0.12.3
4	Channel	Wireless
5	Network Interface	Wireless physical
6	Queue	Drop tail/Priority queue
7	Antenna Type	Omni-Antenna
8	Simulation Time	300 seconds
9	Speed	40 (km/hr)
10	Mac Type	Mac 802_11
11	Interface Queue Length	50
12	Propagation Model	Two ray Propagation
13	Routing Protocols	AODV, Bee Swarm routing protocol, Modified hybrid routing protocol
14	Number of Nodes	40,100
15	Mobility Model	Realistic Mobility Model

5. SIMULATION RESULTS AND ANALYSIS

In this paper, we basically generated Realistic Mobility Model and presented the analysis of the proposed hybrid routing protocol using the performance metrics such as Packet delivery ratio, end-to-end delay and normalized overhead ratio.

For the research of the proposed system we generated realistic mobility model by using SUMO and MOVE. To generate Realistic model we have to create a road map of the research area and its vehicle movement. These model screenshots are shows up in the given below figures:

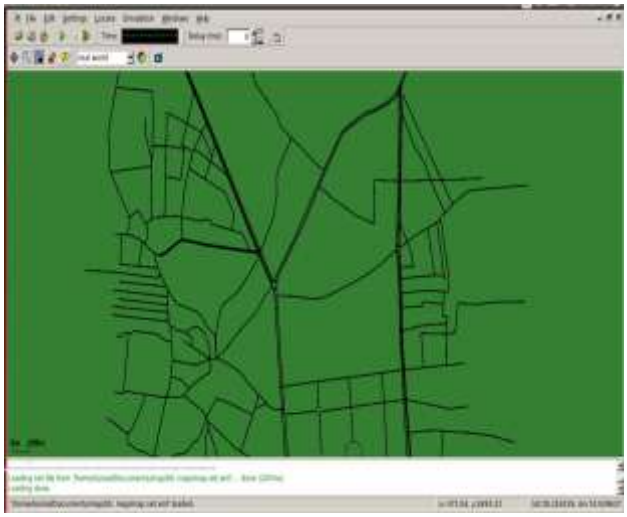


Figure.3 Road Map visualization

The movement of vehicles can be generated automatically or manually by using the vehicle movement editor and we can also specified distinct vehicle route properties by movement editor.

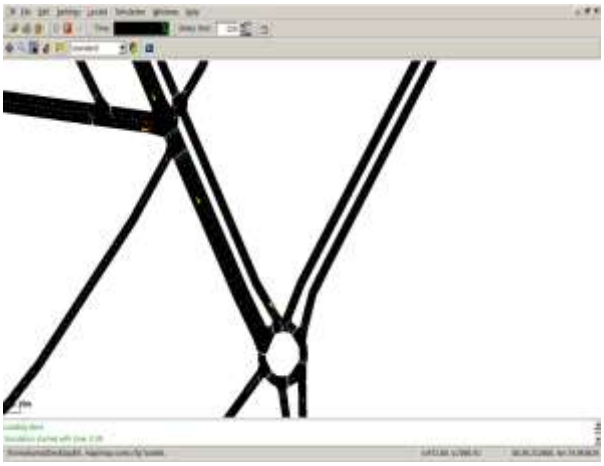


Figure.4 Vehicle Movement on road

5.1 NETWORK SIMULATIONS

The screenshots of simulation in Ns-2 nam trace file shown below in the figure: 5-6 which consists 40 and 100 nodes in the network. In this it captured the node movement and communication between other active vehicles.

For 40 Nodes

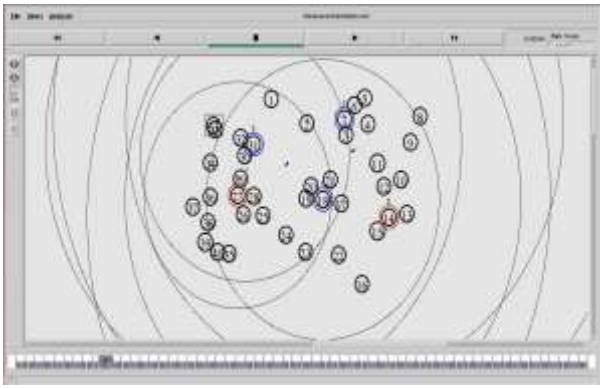


Figure.5 Network scenario for 40 nodes of the vehicular network with random distribution

For 100 Nodes

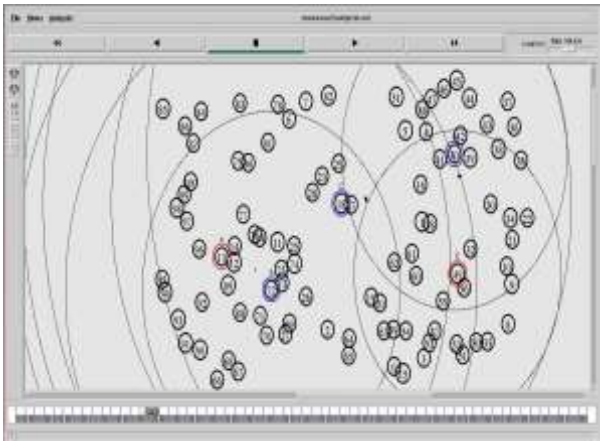


Figure.6 Network scenario for 100 nodes corresponding to the high traffic density

5.2 PERFORMANCE ANALYSIS

The proposed system is simulated and tested for 40,100 nodes to evaluate the performance parameters. Here we analyzed some of the performance metrics such as Packet Delivery Ratio, End-to-End delay, Normalized Overhead Ratio which are shown in the following line graphs.

5.2.1 PACKET DELIVERY RATIO

The Packet delivery ratio in this simulation process is defined as the ratio of the number of data packet s sent to the destination by constant bit rate sources. This illustrates the level of delivered data to the destination.

$$\sum \text{Number of packet receive} / \sum \text{Number of packet send}$$

For 40 Nodes

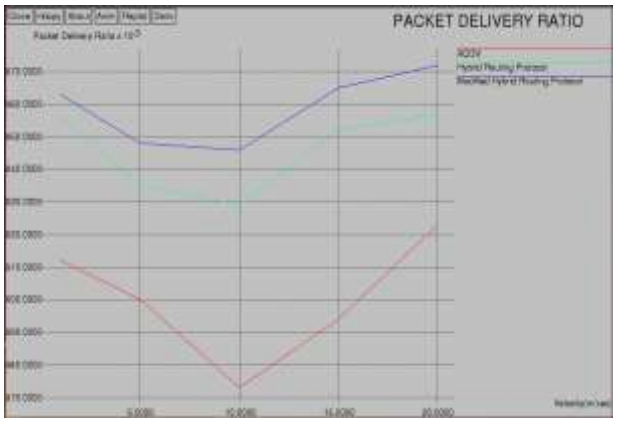


Figure.7 The packet Delivery ratio comparison among the three routing protocols for 40 nodes of the network.

For 100 Nodes

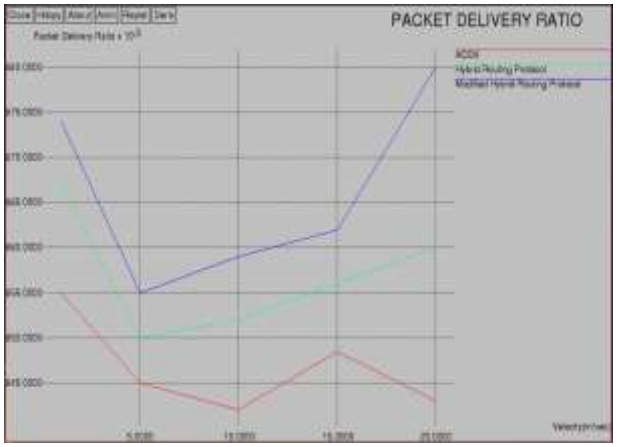


Figure8. The packet Delivery ratio comparison among the three routing protocols for 100 nodes of the network

5.2.2 END-TO-END Delay

End to End delay is one of the important metrics because VANET needs a small latency to transmit data packets quickly. It gives the average overall delay time taken by a packet to travel from source to destination. Delay depends on number of hops and congestion on the network.

Only the data packets that successfully delivered to destination that counted.

∑ (arrive time – send time) / ∑ Number of connections

For 40 Nodes

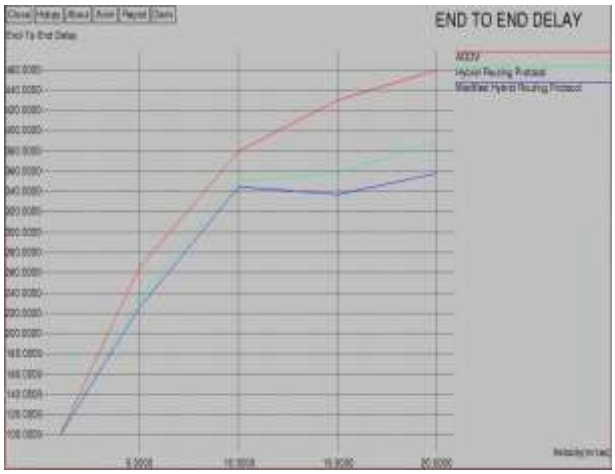


Figure.9 The end to end delay for the proposed algorithm, AODV Hybrid routing protocol for 40 nodes of the network.

For 100 Nodes

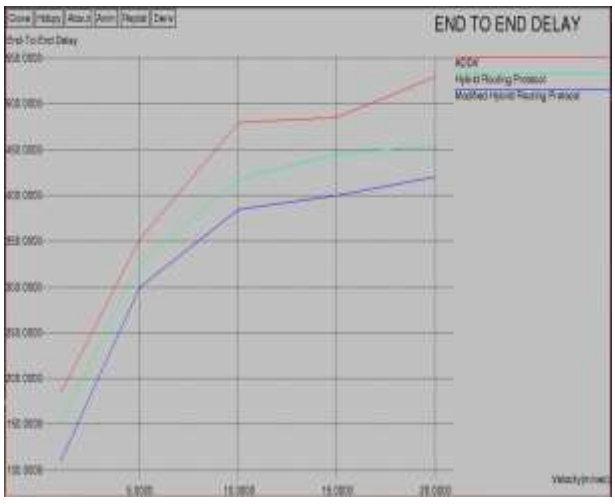


Figure.10 The end to end delay for the proposed algorithm, AODV Hybrid routing protocol for 100 nodes of the network.

5.2.3 NORMALIZED OVERHEAD RATIO :

The Normalized Overhead Ratio is defined as the number of routing packets required for the communication that is the total number of routing packets divided by total number of data packets delivered. This ratio is the measure of the extra packets other than the data packets generated as the percentage of the total data packets.

For 40 Nodes

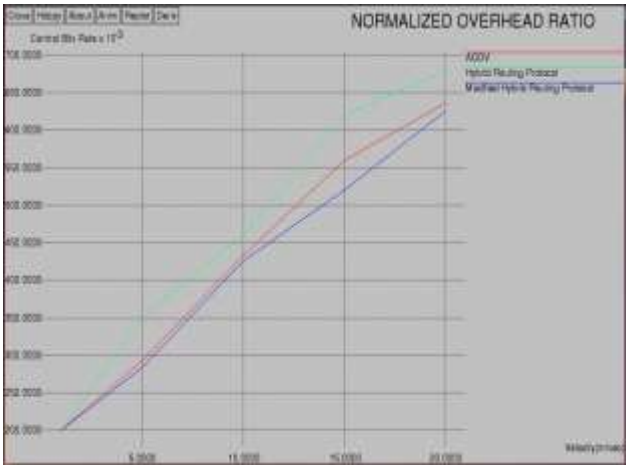


Figure.11 The Normalized overhead ratio of the three routing protocols for 40 nodes

For 100 Nodes

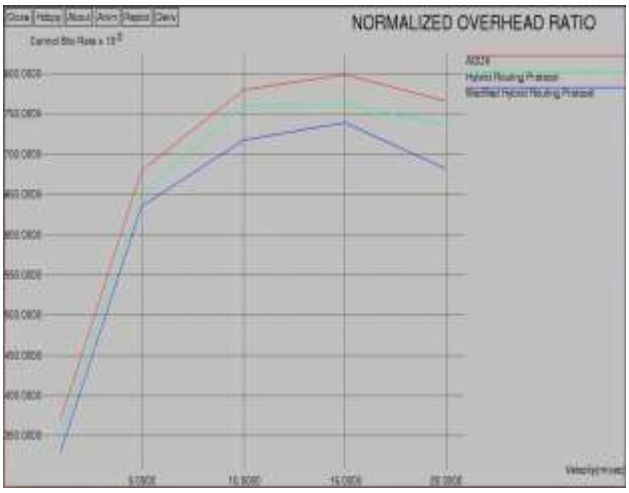


Figure.12 Normalized overhead ratio of the given routing protocols at different velocity intervals for 100 nodes of the network.

6. CONCLUSION

In this paper, the analysis of the proposed Hybrid algorithms is done in realistic scenario of VANET. The Realistic simulation environment is generated using the SUMO and MOVE software and the outcome of this simulation is used for network simulation using NS2. Here on comparing the performance of the introduced protocol with the Bee Swarm Routing Protocol and other static routing protocol AODV in real deployment it shows better performance and efficient results on the basis of the network parameters: end to end delay, packet delivery ratio and normalized overhead ratio.

7. ACKNOWLEDGMENT

I would like to express my sincere gratitude to my advisor Er.Vinod Kumar for his valuable and constructive suggestions during the planning and development of this research work. His guidance helped me in all the time of research and writing of the thesis. Working under him was an extremely knowledgeable experience for me.

8. REFERENCES

[1] Siddhant Jaiswal, Dr D.S Adane, "Hybrid Approach for Routing in Vehicular Ad-hoc Network (VANET) Using Clustering Approach" International Journal of Innovative Research in Computer and Communication Engineering, Vol. 1, Issue 5, July 2013.

[2] V.S.DHAKA,RAMESH C.POONIA and LINESH RAJA, " The Realistic Mobility Evaluation of Vehicular Ad-hoc Network for Indian Automotive Networks" International Journal of Ad hoc, Sensor & Ubiquitous Computing (IJASUC) Vol.5, No.2, April 2014.

[3] Md. Humayun Kabir, "Research Issues on Vehicular Ad hoc Network" International Journal of Engineering Trends and Technology (IJETT) – Volume 6 Number 4- Dec 2013.

[4] Wiseborn Manfe Danquah,D. Turgay Altilar, "HYBRIST Mobility Model - A Novel Hybrid Mobility Model for VANET Simulations" International Journal of Computer Applications (0975 – 8887) ,Volume 86 – No 14, January 2014.

[5] Krishna Kumar Jha, Daya Sagar Baral, " VANET Analysis for Real Time Traffic of Nepal Using SUMO and NS3 under different protocol" Department of Electronics & Computer Engineering, Pulchowk Campus, Institute of Engineering, Tribhuvan University, Nepal.

[6] Feliz Kristianto Karnadi,Zhi Hai Mo,Kun-chan Lan, " Rapid Generation of Realistic Mobility Models for VANET" School of Computer Science and Engineering, University of New South Wales, Sydney, NSW 2052, Australia.

[7] Nataraj B, Dr. T. Kantharaju, " Realistic Mobility Model And Co - Operative Peer To Peer Data Transmission For VANET's Using SUMO And MOVE" IJREAT International Journal of Research in Engineering & Advanced Technology, Volume 3, Issue 6, Dec -Jan, 2016.

[8] Michael Behrisch, Laura Bieker, Jakob Erdmann, Daniel Krajzewicz, " SUMO – Simulation of Urban MObility"Institute of Transportation Systems ,German Aerospace Center ,Rutherfordstr. 2, 12489 Berlin, Germany.

[9] http://www.isi.edu/nsnam/ns/

[10] Kun-chan Lan and Chien-Ming Chou, "Realistic Mobility Models for Vehicular Ad hoc Network (VANET) Simulations" National Cheng Kung University ,Department of Computer Science and Information Engineering ,Tainan, Taiwan.

[11] M.Sumathi, "Swarm Intelligence based Bee inspired routing protocol for multipath routing in MANET" International Journal of Advanced Research in Computer and Communication Engineering , Vol. 4, Issue 2, February 2015.

[12] K.D. Kalambe , A.R. Deshmukh, S.S Dorle, " Particle Swarm Optimization based Routing Protocol for vehicular Ad hoc Network" International Journal of Engineering Research and General Science, ISSN 2091-2730, Volume 3, Issue 1, January-February, 2015 .

[13] Saurabh D. Patil,D.V. Thombare,Vaishali D. Khairnar, " DEMO: Simulation of Realistic Mobility Model and Implementation of 802.11p (DSRC) for Vehicular Networks (VANET)" International Journal of Computer Applications (0975 – 8887) Volume 43– No.21, April 2012.

[14] R.S.Raw, Vikas Toor, and N. Singh, "P ATH DURATION ANALYSIS IN VEHICULAR AD HOC NETWORK" International Journal on Ad-Hoc Networking Systems (IJANS) Vol. 2, No. 4, October 2012.

[15] Adil Baykasoùlu, Lale Özbakır and Pınar Tapkan, "Artificial Bee Colony Algorithm and Its Application to Generalized Assignment Problem" Department of Industrial Engineering, Erciyes University, Turkey.

[16] Pratik Gujar, "BIO-INSPIRED ROUTING PROTOCOL FOR VEHICULAR ADHOC NETWORKS" International Journal of Advancements in Research & Technology, ISSN 2278-7763, Volume 3, Issue 4,April-2014 .

[17] YUN- WEI LIN, YUH- SHYAN CHEN AND SING-LING LEE, "Routing Protocols in Vehicular Ad Hoc Networks: A Survey and Future Perspectives" Department of Computer Science and Information Engineering, National Chung Cheng University, Chia-Yi, Taiwan, R.O.C.

[18] Chitraxi Raj,Urvik Upadhayaya,Twinkle Makwana,Payal Mahida, "Simulation of VANET Using NS-3 and SUMO" International Journal of Advanced Research in Computer Science and Software Engineering, Volume 4, Issue 4, April 2014.

International Journal of Computer Applications Technology and Research
Volume 5– Issue 6, 391 - 394, 2016, ISSN:- 2319–8656

A New Method for Encrypting Digital Data Using Symmetric Key in Information Exchange Spaces

Hadi Hajilar Lahrod
Department of Electrical Engineering, Ahar Branch, Islamic Azad University, Ahar, Iran.

Ali Rahnemaei
Department of Electrical Engineering, Ardabil Branch, Islamic Azad University, Ardabil, Iran.

Seyed Hadi Seyed Hatami
Department of Electrical Engineering, Ardabil Branch, Islamic Azad University, Ardabil, Iran.

Abstract: with the arrival of the information age and much more important information systems and communication in human everyday life, necessity immunization information and communication strategy were also raised. The easiest way to meet this necessity is conventional encryption algorithms. Encryption is a right tool for data protection in an unsecure channel. To this end, from two-method symmetric key encryption and public-key cryptography are used. In this paper we examine text cryptography, one of the most important topics in cryptography. A unique attribute of this kind of encryption has been of interest to many researchers in this field. This paper, considering the symmetric encryption algorithm, provides a text encryption algorithm using a 128-bit key. The proposed algorithm uses a 128-bit key, the text data using the XOR operator to convert the encrypted information. Therefore, the aim of this method is to provide a convenient method for symmetrically encrypting data not to be easily decoded, and finally, the results of the tests show that the proposed method is better in terms of security and speed of execution.

Keywords: Encryption; symmetric; XOR operator; text information; key.

1. INTRODUCTION

One way to provide safety information is encryption. With encryption, confidentiality and message authenticity are preserved. The main problem in cryptography is that the threats neither cannot obtain the original text of encrypted text, nor cannot find the decoding converters even by accessing the original text. In this case, the amount information obtained from the encrypted text and encoding method are important [1]. Cryptography includes of two major components, an algorithm and a key. Algorithm is a converter or mathematical formula. There are a few powerful algorithms that most of them have been published as standards or mathematical papers. Key is a string of binary digits (ones and zeros) which is meaningless by itself. Modern cryptography assumes that an algorithm is known or can be discovered. Key must be kept secret and changed in any stage of implementing. Decryption may use the same pair of algorithms and key. Data encryption algorithms are generally divided into two categories. The first batch contains symmetric encryption algorithms while the second category contains asymmetric cipher algorithms. Asymmetric key encryption algorithms use different keys for encryption and decryption. Many systems allow the one of publicly keys to be released while another private key is kept by its owner. Sender of the message codes the text with the recipient's public key and receptor encrypts it with their private key. In other words, only recipient's private key makes it possible to turn the coded text to the original one. It means that even if the sender accesses the main content text, they cannot achieve to the original text via cipher text. Therefore the coded message will be meaningless for any recipient rather than the real one [2]. Therefore, with studying algorithms that has been used before in this context, a new solution for encrypting confidential information can affect and help establishment of security in communications. In this paper we propose a new method for encrypting digital data using asymmetric key in the exchange of information spaces.

2. SECURITY REQUIREMENTS

Each encryption algorithm that is used to create data security should ensure some specific security requirements which are as follows:

2.1 Confidentiality of data's

The main purpose in security issues is to maintain data confidentially from unauthorized user accesses. In fact, data confidentiality is means maintenance of information secretly. Different protocols use various cryptographic methods.

2.2 Data integrity

Making information confidential, external factors cannot steal information, but it does not meant the secure and healthy data. Data integrity ensures that the received data has been remained unchanged during transmitting.

2.3 Data Novelty

Data novelty ensures that received data is new and is not a repeat of previous messages.

2.4 Authentication

Encryption algorithms must have the authentication ability of the received data, and can ensure the accuracy of the transmitter. The sender of information cannot, in the future, deny sending or its provisions. Various security protocols use different encryption algorithms to meet these requirements.

3. CLASSIFICATION OF ENCRYPTION METHODS

An encryption algorithm is a set of rules and mathematical relationships which ends in a difference and clutter in the data. Modern encryption algorithms can be classified based on two criteria; functional keys and type of input. Data encryption method is divided, based on input data type, into two blocks and streams. Most of the techniques presented are based on the image encryption [3][4].

4. SYMMETRIC ENCRYPTION

In general, symmetric encryption has two parts, sender and receiver, which want to communicate with each other in an insecure channel, without letting someone else to get any information on the relationship. The purpose of encryption algorithms is to protect the security of a message that is transmitted through an insecure channel. In general an organization symmetric encryption has two functions, encryption E and decryption D. The Encryption function E receives the inverse of the original text as an input, and transforms it under the key K to cipher text$C = E_k(P)$. The encryption function $D = E^{-1}$ is reverse function of encryption function, so $P = D_K(C)$. the key B is also called secret key or shared key. In Fig.1, a general block diagram of symmetric encryption system has been shown.

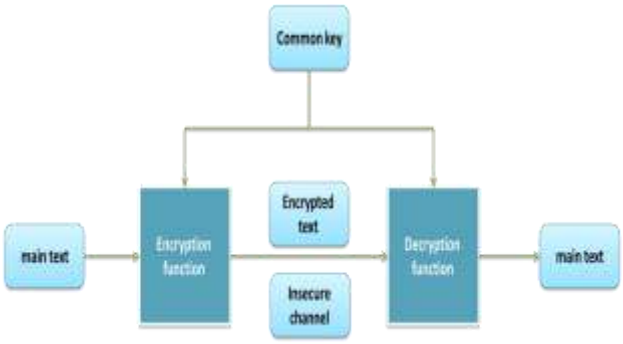


Fig.1 Block diagram of symmetric encryption system

5. RELATED WORKS

Security information in communication technology is one of the issues that has occupied scholars of this area, because with compromising information security, serious problems happen to come for the message and threaten its integrity. Due to this problem, scientists went on to make secret before exchanging information and in the consequent, they reduce or neutralize the effect of inhibiting factors. With technological advances and the advent of modern communications, the need to provide new methods in this area is greater than ever. Thus, according to expand of the use of communication technology in transmitting critical data, the privacy and data security are of utmost importance. There have been done multiple algorithms to encrypt data too [5][6][7]. In some other activities encryption of data has been investigated for various methods [8][9]. Various proposed methods in encryption information are divided generally into two categories: stream encryption and blocks encryption [10][11]. In the first kind, at any moment, encryption is done bit by bit or character by character, but in the second the entire string is encrypted and transmitted at once. Generally, for both encryption methods, the keys are used which are made of pseudorandom numbers. There are several methods for generating random numbers such as modular arithmetic generators, linear and non-linear, linear recurrence registers (LFSR) and a non-linear Cellular Automata (CA) and so on [12]. Access to information stored in computer databases has greatly increased. Much of the information stored is highly confidential, which is not visible to the public. Data security is studied using encryption techniques research conducted in an article with this theme, a new encryption algorithm is provided based on concept of block encryption, and logical operations such as XOR and shift operations are used as well. The proposed algorithm improves encryption security by symmetric putting layers [13]. Visual secret sharing (VAS) is a visual encryption scheme which decodes secret messages to several big stocks [14].

6. ENCRYPTION APPROACH

The proposed encryption system is one of the symmetric systems that guarantee an absolute security. Overall approach of this system is to generate key length of plain text and XOR keys with text. In this method, the receiver of the message must have the encryption key up with XOR again, restore the original text. Totally randomness key, full security for the proposed system ensures this approach also high computational security in the face of the enemy. In building public key cryptography systems, objective of securing is computational. In this case, although these systems are not secure against invaders, such strikers do not exist in the real world.

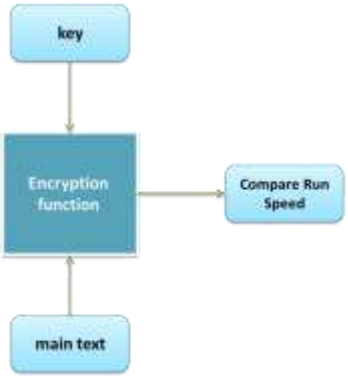


Fig. 2 Main operation the proposed algorithm

7. DESCRIPTION OF THE PROPOSED ALGORITHM

In some methods, particular keys are used to perform encryption and decryption. Used keys can be the same in encryption and decryption and/or are used differently in these two factors. There are other algorithms that do not use keys for encryption. These methods create kind of clutter in messages and through this prevent the discovery of information, by unauthorized persons. Different encryption algorithms, as symmetric, asymmetric and hash functions are used for hiding and for each of these algorithms there have been proposed various methods. In the proposed method, we consider a text file, according to the same text we convert it to an array of characters. Then these characters are converted to an array of bits. Key Length for encrypt information should be ideal, if key length is greater, the security is then better and speed is lower, and vice versa, if the key length is less, security is less and encryption speed increases. In this paper we consider 128-bit key length. After selecting key length, it will be applied on the bits array in the range of 2/128. We divide bits array to 128 bits, this is our method for performing encryption. In each stage, the XOR operation will be also applied; XOR operation is as follows: after division of bits array to 128 bits, the first part of the array of bits that is 128-bit with key is done by XOR. After encrypting the first part of the array, this time half a second part of the array with half of the first part of the array with XOR action key is done; we do this action until completing the array and cryptographic operation. In the proposed method we used XOR operator to easily perform the encryption and decryption. How to do encrypt text is shown in Fig. 3:

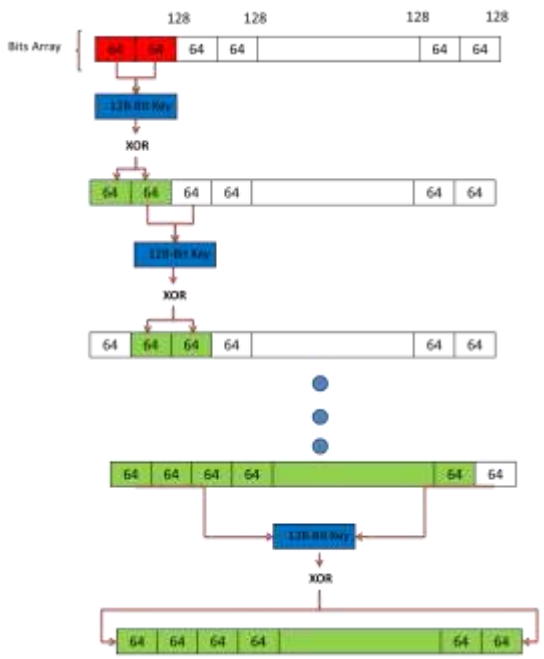


Fig.3 Description of the proposed approach

Encrypting the information
Data encryption algorithm in this method is very simple.

1. Converting secret text information to array of characters
2. Considering key length 128 bits
3. Key actions on array with range of 2/128
4. XOR perform on array with key consideration

8. IMPLEMENTATION RESULTS

For performing the test we used different texts for encryption. According to the results, it can be observed that the proposed algorithm is the fastest algorithm based on run-time. These results have been achieved using MATLAB 2009 software. The study of results shows that the encoded information output resulting from encryption method has been reported with security and high quality. We used "English" text in implementation of the proposed algorithm for encryption. In Fig.4 the English confidential text is used for encryption. Also Fig.5 and 6 show, respectively, English text information before the operation encryption and information encrypted after encryption. So after encryption operation and applying the proposed method, confidential data turns out to be in the shape of Fig.7.

Fig.4 Confidential information for encrypting

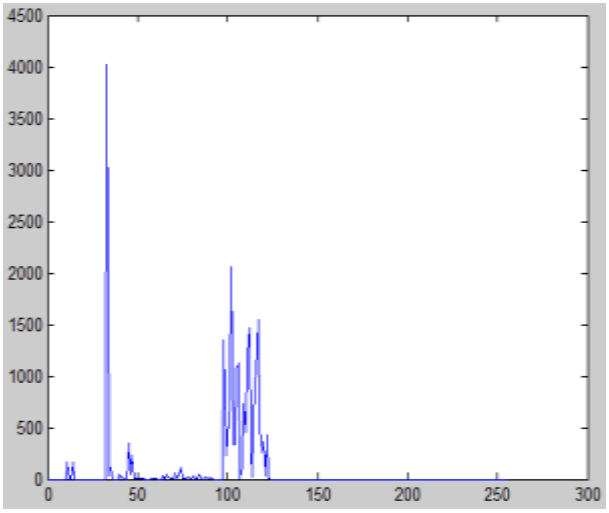


Fig.5 Text information for encryption

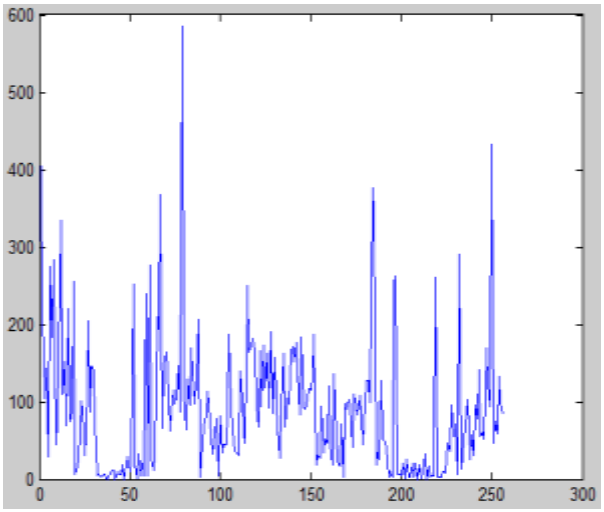


Fig.6 information encrypted

Fig.7 information encrypted

In comparison between different encryption algorithms with data size and various keys size, the proposed algorithm represents the best performance. Table 1 compares the processing speed of the proposed encryption algorithm with

different key lengths with three known block ciphers. As the table shows the proposed algorithm, with length 128-bit, runs faster than other encryption methods and can serve up operations. The proposed algorithm, due to speed and simpler being implemented, is a type of symmetric cryptography for protecting information in network communication and can be used for protection of data on public channels. Time speed of encrypting and decrypting depends largely on encryption key length and file size. For obtaining confidential information and decoding it, decryption key is generated as the same generation method of encryption key, and output encrypted text being XOR to obtain a plain text.

Table 1. Comparison symmetric ciphers processing speed with the proposed algorithm

Length key	Type code	(Mbps) speed
56	DES	9
168	3DES	3
Variable	RC2	0.9
Variable	RC4	45
128	Proposed method	52

9. CONCLUSION

This article provides a secure method for encrypting text with respect to the symmetric key and XOR operator. The proposed method converts confidential information into an array of characters, then using a 128-bit symmetric key encrypts information via XOR operator, In fact the main stage is to use key and to take the XOR operator. For decryption we act in contrast to encryption. In the decode step, in first, we have a symmetric key along with the encrypted information. And according to encryption routine, we select the text again character by character and, using XOR operator, decrypt confidential information. Evaluation of the proposed method showed that encryption of confidential information is better than other methods. Security and data confidentiality in the proposed algorithm is maximum. After implementation and testing various data on the proposed algorithm we reached the conclusion that our proposed algorithm, compared to algorithms provided in the field, is more efficient and decryption of it, without the key, is more difficult and even impossible to solve. Also the algorithm speed, due to the simplicity of the method, is of high operating and computing speed. As suggestions for future studies we can note to examine asymmetric algorithms and data hiding by the algorithm.

10. REFERENCES

[1] A. Menezes, P.van Oorschot and S. Vanstone, "Handbook of Applied Cryptography", CRC Press, 1996.

[2] Jacob John, “Cryptography for resource constrained devices :A Survey”, Article in a regular journal, International Journal on Computer Science & Engineering, 2012.

[3] Ching-Nung Yang, Pin-Wei Chen, Hsiang-Wen Shih, Cheonshik Kim, Aspect ratio invariant visual cryptography by image filtering and resizing, Personal Ubiquitous Comput. 17 (5) (2013) 843–850.

[4] Kai-Hui Lee, Pei-Ling Chiu, Image size invariant visual cryptography for general access structures subject to display quality constraints, IEEE Trans. Image Process. 22 (10) (2013) 3830–3841.

[5] Ching-Nung Yang, Ting-Hao Chung, " A general multi-secret visual cryptography scheme", Elsevier, Optics Communications 283 (2010) 4949–4962.

[6] Jasdeep Singh Bhalla, PreetiNagrath ,Article in a regular journal, International Journal of Scientific & Research publications,”Nested Digital Image Watermarking Technique Using Blowfish Encryption Algorithm”, Volume 3, Issue 4, April 2013.

[7] A.Nath, S.Ghosh, M.A.Mallik, "Symmetric key cryptography using random key generator", Proceedings of International conference on SAM-2010 held at Las Vegas (USA) 12-15 July, 2010, Vol-2, P-239-244.

[8] Sui L, Duan K, Liang J, "Double-image encryption based on discrete multiple-parameter fractional angular transform and two-coupled logistic maps", Optics Communications Volume343, Elsevier ,2015, PP. 140–149.

[9] Sajasi S, Eftekhari Moghadam A.M, "An adaptive image steganographic scheme based on Noise Visibility Function and an optimal chaotic based encryption method", Applied Soft Computing, Volume 30, Elsevier, May 2015, PP. 375–389.

[10] Nikita Arora,Yogita Gigras, "Block and Stream Cipher Based Cryptographic Algorithms: A Survey", International Journal of Information and Computation Technology, ISSN 0974-2239 Volume 4, Number 2 (2014), pp. 189-196.

[11] D.J.C. MacKay, "Information Theory, Inference, andLearning Algorithms", June 26, 2003, Cambridge university Press.

[12] S. Haykin, "Neural Networks – A Comprehensive Fundation", 2th Edition, Prentice Hall, 1999.

[13] Vishwa gupta, Gajendra Singh, Ravindra Gupta, "Advance cryptography algorithm for improving data security", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 2, Issue 1, January 2012.

[14] Lee, CH, Chen H, Liu H, Chen G, "A new visual cryptography with multi-level encoding", Journal of Visual Languages and Computing, Elsevier, Volume 25, Issue 3, June 2014, PP 243–250.

International Journal of Computer Applications Technology and Research
Volume 5– Issue 6, 395 - 402, 2016, ISSN:- 2319–8656

A Review on Feature Selection Methods For Classification Tasks

Mary Walowe Mwadulo
Department of Information Technology, Meru University of Science and Technology,
P.O BOX 972-60200 Meru, Kenya.

Abstract: In recent years, application of feature selection methods in medical datasets has greatly increased. The challenging task in feature selection is how to obtain an optimal subset of relevant and non redundant features which will give an optimal solution without increasing the complexity of the modeling task. Thus, there is a need to make practitioners aware of feature selection methods that have been successfully applied in medical data sets and highlight future trends in this area. The findings indicate that most existing feature selection methods depend on univariate ranking that does not take into account interactions between variables, overlook stability of the selection algorithms and the methods that produce good accuracy employ more number of features. However, developing a universal method that achieves the best classification accuracy with fewer features is still an open research area.

Keywords: Feature selection, attribute, dimensionality reduction, optimal subset, classification

1. INTRODUCTION

In recent years, the need to apply feature selection methods in medical datasets has greatly increased. This is because most medical datasets have large number of samples of high dimensional features. This makes it impractical, computationally expensive and causes reduction in classification accuracy when an entire input set is used. Thus, there is a need to reduce the number of features to manageable sizes which can be achieved through feature selection. Choosing an appropriate feature selection method is a non-trivial task, thus the motivation of this review is to make practitioners aware of feature selection methods that have been successfully applied in medical data sets and highlight future trends in this area.

A feature is a distinctive attribute that can be used to measure a process under observation [1]. Feature selection is a dimensionality reduction technique that reduces the number of attributes to a manageable size for processing and analysis [1]. In contrast to other dimensionality reduction techniques, feature selection does not alter the original feature set rather selects a subset by eliminating all the features whose presence in the dataset does not positively affect the learning model [1]. Thus preserves the original semantics of the features which makes it easy to interpret. Using a set of features a machine learning technique can perform classification. Classification is a machine learning task that involves assigning known class labels to training data [2].

The set of features used in model construction in the only source of information for any learning algorithm, thus it is extremely important to select an optimal subset that will be a representative of the original set. Selecting an optimal subset of relevant and non redundant features is a challenging task. Since there is a trend off, if too many features are selected it causes the classifier to have a high workload which can decrease the classification accuracy. On the other hand, if too few features are selected there is a possibility of eliminating features that would have increased the classification accuracy. Thus, there is a need to get an optimal subset of relevant and non redundant features which will give an optimal solution without decreasing the classification accuracy. No known effective method has been devised to select an optimal subset.

Feature selection helps in understanding data, reducing computational requirements, reducing the curse of dimensionality and improving the prediction performance [1]. By combining several feature selection methods, the curse of dimensionality can be reduced and classification accuracy of modeling tasks improved.

The remaining part of this paper is structured as follows: section 2 presents Feature selection techniques, section 3 presents the discussion and section 4 presents conclusion.

2. FEATURE SELECTION TECHNIQUES

Feature selection is a pre processing technique used in machine learning to remove irrelevant and redundant attributes for the purpose of increasing learning accuracy [1]. Feature selection does not only imply to cardinality reduction (imposing an arbitrary or predefined cutoff on the number of attributes that can be considered when building a model) but also the choice of attributes which could be based on presence or lack of interaction among the attributes and the classification algorithm. This means that the modeling tool actively selects or discards attributes based on their usefulness for analysis. Feature selection is necessary because the high dimensionality and vast amount of data poses a challenge to the learning task. In the presence of many irrelevant features some of which do not add much value during the learning process, learning models tend to become computationally complex, over fit, become less comprehensible and decrease learning accuracy. Feature selection is one effective way to identify relevant features for dimensionality reduction. However, the advantages of feature selection come with extra effort of trying to get an optimal subset that will be a true representation of the original dataset. In the context of classification, feature selection techniques can be categorized into Filter methods, wrapper methods, embedded methods and hybrid methods.

2.1 Filter methods

Filter methods are feature ranking techniques that evaluate the relevance of features by looking at the intrinsic properties of the data independent of the classification algorithm [2], [3], [4]. A suitable ranking criterion is used to score the variables and a threshold is used to remove the variable below the threshold [1]. Afterwards this subset of features is used as input to the classification algorithm. Filter methods assess the relevance of features using measures like distance, information, correlation and consistency [5]. Advantages of filter methods are that they are fast, scalable and independent of a learning algorithm. As a result feature selection needs to be performed only once, and then different classifiers can be evaluated [2]. Disadvantages of filter techniques is that they lack interaction with the classifier which makes them generate general results and lower classification accuracy [2], [6]. Filter methods can be categorized into univariate and multivariate. Univariate filter methods ignore feature dependencies which can lead to selection of redundant features and worst classification performance when compared to other feature selection techniques [2]. On the other hand, multivariate filter methods model feature dependencies independent of the classifier. In addition to evaluating class relevance like univariate, they also calculate the dependency between each feature pair [2]. Some univariate filter feature selection methods include:

2.1.1 Information gain (IG)

It is a symmetrical measure of dependency between two variables. The information gained about Y after observing X is equal to the information gained about X after observing Y [6]. Selects candidate features with more information, for each feature a score is obtained based on how much more information about the class is gained when using that feature. The level of features usefulness is determined by how great is the decrease in entropy of the class when considered with the corresponding features individually [3]. Disadvantage of IG is that it favors features with more values even when they may not be more informative [3].

IG is defined as:

$$IG(X; Y) = H(X) - H(X|Y)$$

2.1.2 Gain Ratio (GR)

The Gain Ratio is a non-symmetrical measure that is introduced to compensate for the bias of the IG [6]. GR is given by

$$GR = \frac{IG}{H(X)}$$

When the variable Y has to be predicted, we normalize the IG by dividing by the entropy of X, and vice versa. Due to this normalization, the GR values always fall in the range [0, 1]. A value of GR = 1 indicates that the knowledge of X completely predicts Y, and GR = 0 means that there is no relation between Y and X. In opposition to IG, the GR favors variables with fewer values [6].

2.1.3 Symmetric Uncertainty (SU)

This is a correlation measure between the features and the target class. The Symmetrical Uncertainty criterion compensates for the inherent bias of IG by dividing it by the sum of the entropies of X and Y [6]. Features with a high Symmetric Uncertainty value get a higher value.

SU takes values, which are normalized to the range [0, 1] because of the correction factor 2. A value of SU = 1 means that the knowledge of one feature completely predicts, and the other SU = 0 indicates, that X and Yare uncorrelated [6]. A weakness of SU is that it is biased towards features with fewer values [3]. It is a normalized information theoretic measure which uses entropy and conditional entropy values to calculate dependencies of features [7].

SU is defined as:

$$SU\,(X,Y) = \frac{2\,IG(X;Y)}{H(X)+H(Y)}$$

Multivariate filter techniques which incorporate a degree of feature dependencies that can be used to solve the problem. Some of multivariate filter methods include:

2.1.4 Correlation based Feature Selection (CFS)

Correlation based feature selection method evaluate subsets of features by selecting feature subsets contain features highly correlated with the classification, yet uncorrelated to each other. CFS evaluates a subset by considering the predictive ability of each one of its features individually and also their degree of redundancy (or correlation). This means that given a function, the algorithm can decide on its next moves by selecting the option that maximizes the output of this function [6].

2.1.5 Markov blanket Filter (MBF)

Markov blanket Filter method finds features that are independent of the class label so that removing them will not affect the accuracy [6]. It does not require one to specify a variable ordering, nor to fix an upper bound on the number of parents allowed for each node, and this makes MBF both more general and more appealing for application to domains where no prior knowledge can be used to constrain the learning process [7].

2.1.6 Fast Correlation based Feature Selection (FCBF)

It is a feature selection method which starts with full set of features, uses symmetrical uncertainty to calculate dependences of features and finds best subset using backward selection technique with sequential search strategy. It has an inside stopping criterion that makes it stop when there are no features left to eliminate. It is a correlation based feature subset selection method which runs, in general, significantly faster than other subset selection methods [7].

2.1.7 Minimum Redundancy Maximum Relevance (MRMR)

It is a multivariate feature selection method which maximizes the relevancy of features with the class label while it minimizes the redundancy in each class [6]. It starts with an empty set, uses mutual information (a symmetrical information theoretic measure that measures the amount of information that can be obtained about one random variable by observing another) to weight features and forward selection technique with sequential search strategy to find the best subset of features. It has a parameter k which enables it to stop when there are k features in the selected feature subset [7]. MRMR does not deal with the type of dependency rather the quantity of dependency (it uses mutual information) which can lead to inaccurate ordering of the variables.

Pandey etl. [1], [8], Used information gain for feature selection which showed remarkable result. In [7] Modified Fast Correlation Feature selection method by giving every feature a temporary

predominance in the elimination process and making them start eliminating features from the features which are least correlated with the class. An iteration process that allows one feature to eliminate one feature per iteration which makes elimination process more balanced.

In [2], adopted a two phase feature selection method, where in the first phase, they combined information gain and symmetric uncertainty to generate two subsets of reliable features. In the second phase, the two subsets are merged, weighted and ranked to extract the most important features. Combination of two filtering methods, lead to higher accuracy of intrusion detection. In [5] a four stage Multi Filtration Feature Selection (MFFS) method was introduced. The method adjusts variance coverage and builds the model with the value at which maximum classification accuracy is obtained. In stage one, relevant features are generated using Principal Component Analysis (PCA), stage two, features are ranked using correlation feature selection which is improved by employing symmetric uncertainty in stage three. Finally the system is validated against standard classifier models. The results showed that classification accuracy based on the selected subset by Multi Filtration Feature Selection (MFFS) method was better than that based on the original feature set. Authors in [6] devised a three stage hybrid feature selection approach, that recommended selecting features at the intersection of information gain and Significance analysis of Micro array (SAM). The intersection features are then subjected to mRMR to minimize redundancy in the second stage. Finally, Support Vector Machine Recursive Feature Elimination (SVM-RFE) is applied to choose the most discriminate genes.

Karimi etal. [3], utilized both feature space and sample domain in two phases. The first phase filters and resample the sample domain and the second phase adopted a hybrid procedure by information gain, wrapper subset evaluation and genetic search to find the optimal feature space.

2.2 Wrappers

They use the predictor as a black box and the predictor performance as the objective function to evaluate the variable subset [1], [2],[3]. A search procedure in the space of possible feature subset is defined and various subsets of features are generated and evaluated. The evaluation of a specific subset of features is obtained by training and testing a specific classification model, making this approach tailored to a specific classification algorithm [2].Advantages of this approach is that it includes the interaction between feature subset search and model selection, and the ability to take into account feature dependencies [2].A common drawback is that it has a higher risk of over fitting than filter techniques and are computationally intensive, especially if the building classifier has a high computational cost. Over fitting occurs if the classifier model learns the data too well and provides poor generalization capability.

Wrappers can be categorized into Sequential selection algorithms and Heuristic search algorithms.

2.2.1 Sequential selection algorithms

This algorithm can do forward or backward selection. With Sequential Forward Selection (SFFS) algorithm, you start with an empty set and add one feature for the first step which gives the highest value for the objective function. From the second step onwards the remaining features are added individually to the current subset and the new subset is evaluated. The individual feature is permanently included in the subset if it gives the maximum classification accuracy. The process is repeated until the required number of features is added. This is a naive SFS algorithm since the dependency between the features is not accounted for [1]; this is impractical for feature subset selection from a large number of samples of high dimensionality features [5]. The Sequential Floating Forward Selection (SFFS) [9] algorithm is more flexible than the naive SFS because it introduces an additional backtracking step. The first step of the algorithm is the same as the SFS algorithm which adds one feature at a time based on the objective function. The SFFS algorithm adds another step which excludes one feature at a time from the subset obtained in the first step and evaluates the new subsets. If excluding a feature increases the value of the objective function

then that feature is removed and goes back to the first step with the new reduced subset or else the algorithm is repeated from the top. This process is repeated until the required number of features is added or required performance is reached. The SFS and SFFS methods suffer from producing nested subsets since the forward inclusion was always unconditional which means that two highly correlated variables might be included if it gave the highest performance in the SFS evaluation.

To avoid the nesting effect, adaptive version of the SFFS was developed in [10]. The Adaptive Sequential Forward Floating Selection (ASFFS) algorithm used a parameter r which would specify the number of features to be added in the inclusion phase which was calculated adaptively. The parameter o would be used in the exclusion phase to remove maximum number of features if it increased the performance. The ASFFS attempted to obtain a less redundant subset than the SFFS algorithm.

A different sequential selection approach is Sequential Backward Selection (SBS). It is similar to SFS but the algorithm starts from the complete set of variables and removes one feature at a time whose removal gives the lowest decrease in predictor performance.

2.2.2 Heuristic search algorithms

The heuristic search algorithms evaluate different subsets to optimize the objective function. Different subsets are generated either by searching around in a search space or by generating solutions to the optimization problem.

Genetic algorithms (GA) is a general adaptive optimization search method based Darwinian principle of 'survival of the fittest', GA works with a set of candidate solutions called a population and obtains the optimal solution after a series of iterative computations. GA evaluates each individual's fitness, i.e. quality of the solution, through a fitness function. The fitter chromosomes have higher probability to be kept in the next generation or be selected into the recombination pool using the tournament selection methods. If the fittest individual or chromosome in a population cannot meet the requirement, successive populations will be reproduced to provide more alternate solutions. The crossover and mutation functions are the main operators that randomly transform the chromosomes and finally impact their fitness value. The evolution will not stop until acceptable results are obtained. Associated with the characteristics of exploitation and exploration search, GA can deal with large search spaces efficiently, and hence has less chance to get local optimal solution than other algorithms [11]. GAs offer a particularly attractive approach for problems like feature subset selection since they are generally quite effective for rapid global search of large, non-linear and poorly understood spaces. GAs are based on an imitation of the biological process in which new and better populations among different species are developed during evolution [12]. Thus, unlike most standard heuristics, GA uses information of a population (individuals) of solutions when they search for better solutions.

In [13] combined Symmetric Uncertainty and Genetic Algorithm for feature selection based on the Naïve Bayes classifier. Experimental results conducted over several UCI datasets revealed that higher level of dimensionality reduction was achieved by selecting less number of features than other methods.

In [3] propose a framework based on a genetic algorithm (GA) for feature subset selection that combines various existing feature selection methods. The goal is to effectively utilize useful information from different feature selection methods to select better feature subsets with smaller size and/or higher classification performance in comparison with the existing methods. Multiple selection criteria are combined by a genetic algorithm to improve feature subset selection.

In [12] used a preprocessed statistical parametric mapping software and PCA were used for dimension reduction. Then, independent components of the new data (given by PCA) were estimated using Independent Component Analysis (ICA) method. For feature extraction, LBP histogram extraction technique was used for all estimated components. Genetic Algorithm was used for selection of the most significant histogram bins, in next step. Then, linear discriminant analysis (LDA) is performed to further extract features that maximize the ratio of between-class and within-class variability. Finally, a classifier based on Euclidean distance was used for classification. In [4] adopted an oversampling approach in which the minority class is oversampled by creating synthetic examples rather than by

oversampling with replacement. The synthetic examples are generated in a less application specific manner, by operating in feature space rather than sample domain. Selective Bayesian which uses a forward and backward greedy search method is applied to find a feature subset from the whole space of entire features. It uses the accuracy of Naïve Bayes classifier on the training data to evaluate feature subsets, and considers adding each unselected feature which can improve the accuracy on each iteration. Entropy measure is then calculated and used to measure uncertainty of a class attribute using information gain. Genetic algorithm is applied as a function optimizer.

2.3 Embedded methods

Embedded methods interact with learning algorithm at a lower computational cost than the wrapper approach [1]. It captures feature dependencies and considers not only relations between one input features and the output feature, but also searches locally for features that allow better local discrimination. It uses the independent criteria to decide the optimal subset for a know cardinality. The learning algorithm is used to select the final optimal subset among the optimal subsets across different cardinality [3].This approach has the advantage of including the interaction with the classification model , while at the same time being far less computationally intensive than wrapper methods [2].

2.4 Hybrid methods

Hybrid methods are based on sequential approach where the first step is usually based on filter methods to reduce the number of features used in the second stage. Afterwards a wrapper method is employed to select the desired number of features using this reduced set [4].

3. DISCUSSION

Classification accuracy is very important on medical data sets; however, medical data sets have many features which can be irrelevant and redundant. The irrelevant and redundant features can overload the classifier and lead to decreased classification accuracy. Thus, there is need to reduce the input set to manageable sizes. To solve this problem [2], [3], [6] aspired to reduce the number of features before presenting it for classification.

Selecting an appropriate set of features is extremely important since the feature set selected is the only source of information for any learning algorithm using the data of interest. A goal of feature selection is to avoid selecting too many or too few features than is necessary. If too few features are selected, there is a possibility that the information content in this set of features is low, on the other hand, if too many (irrelevant) features are selected, the effects due to noise may overshadow the information present. Hence this is a trade off that must be considered when applying feature selection methods.

Researchers have tried to address the issue of feature subset selection through filter methods [1], [2], [5], [7] some of which provide a ranking criterion. These methods are fast and scalable; however, they ignore feature dependencies and interaction with the classifier. This makes their results unrealistic since a given feature might provide more information when present with certain other features than when considered by itself. Thus, it is important to consider features not only in relation to the class but also in relation to each other. Again features should be selected as a set, rather than selecting the best features to form the (supposedly) best set. The best individual feature does not necessarily constitute the best set of features. However in most real world situations, it is not know what the best set of features is neither the number of features in such a set.

4. CONCLUSION

After reviewing the work on feature selection, it is observed that obtaining an optimal subset of relevant and non redundant features is a non trivial task. Most of the existing methods in the literature depend on univariate ranking that does not take into account interactions between the variables already included in the selected subsets and the remaining one, overlook stability of the selection algorithm and the methods that produce good accuracy employ more number of features which affects the classification accuracy. This paper attempts to reveal that a holistic and Universal method

that achieves the best classification accuracy with fewer features is still an open research area.

5. REFERENCES

[1] Girish Chandrashekar, Ferat Sahin, (2014). "A survey on feature selection methods". Computers and Electrical Engineering.

[2] Yvan Saeys, Inak Inza, Pedro Larranaga, (2007). "A review of Feature Selection techniques in bioinformatics". Bioinformatics, Oxford University press.

[3] Feng Tan, Xuezheng Fu, Yanqing Zhang, Anu G. Bourgeois, (2008). "A genetic algorithm-based method for feature subset selection". Soft Comput.

[4] Muhammad Shakil Pervez, Dewan Md. Farid ,(2015). "Literature Review of Feature Selection for mining Tasks".International Journal of Computer Application, Vol 116, No. 21.

[5] S. Sasikala, S. Appavu alias Balamurugan, S. Geetha, (2014). "Multi Filtration Feature Selection (MFFS) to improve discriminatory ability in clinical data set". Applied Computing and Informatics.

[6] Hall, M. A. & Smith, L. A. (1998). Practical feature subset selection for machine learning. In C. McDonald (Ed.), Computer Science '98 Proceedings of the 21st Australasian Computer Science Conference ACSC'98, Perth.

[7] Baris Senliol, Gokhan Gulgezen, Lei Yu, Zehra Cataltepe, (2008)." Fast Correlation Based Filter (FCBF) with a Different Search Strategy". Computer and Information Science, 23rd international symponsium.

[8] B.Azhagusundari, Antony Selvadoss Thanamani, (2013). "Feature Selection based on Information Gain". International Journal of Innovative Technology and Exploring Engineering (IJITEE), Vol 2, issue 2.

[9] P. Pudil , J. Novovicova , j. Kittler (1994). "Floating search methods in feature selection". Pattern Recognition Letters.

[10] P. Somol, P. Pudil , J. Novovicova, P. Paclik (1999)." Adaptive Floating search methods in feature selection". Pattern Recognition Letters.

[11] Li Zhuo, Jing Zheng, Fang Wang , Xia Li , Bin Ai , Junping Qian, (2008). "A Genetic Algorithm Based Wrapper Feature Selection Method For Classification of Hyperspectral Images Using Support Vector Machine". The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. Vol. 37.

[12] H. Shahamat, A. A. Pouyan, (2014). Feature selection using genetic algorithm for classification of schizophrenia using fMRI data.

[13] Bai-Ning Jiang, Xiang-Qian Ding, Lin-Tao Ma, Ying He, Tao Wang, Wei-Wei Xie, (2008). A Hybrid Feature Selection Algorithm: Combination of Symmetrical Uncertainty and Genetic Algorithms. The Second International Symposium on Optimization and Systems Biology (OSB'08).

[14] Suman Pandey, Anshu Tiwari, Akhilesh Kumar Shirivas ,Vivek Sharma, (2015). "Thyroid Classification using Ensemble Model with feature selection". International Journal of Computer Science and Information Technologies, Vol. 6 (3).

[15] Shailendra Singh, Sanjay Silakari, (2009). "An ensemble approach for feature selection of Cyber Attack Dataset". International Journal of Computer Science and Information Security, Vol 6, No. 2.

[16] Zahra karimi, Mohammad Mansour, Ali Harounabadi, (2013). " Feature Ranking in Intrusion Detection Dataset using Combination of Filtering Methods". International Journal of Computer Applications. Vol. 78. No. 4.

[17] Mehdi Naseriparsa, Amir-Masoud Bidgoli, Touraj Varaee, (2013). "A Hybrid Feature Selection Method to Improve Performance of a Group of Classification Algorithms". International Journal of Computer Applications .Vol. 69 No.17.

[18] Ahmed Soufi Abou-Taleb, Ahmed Ahmed Mohamed, Osama Abdo Mohamed, Amr Hassan Abedelhalim,

(2013). “Hybridizing Filters and Wrapper Approaches for Improving the Classification Accuracy of Microarray Dataset”. International Journal of Soft Computing and Engineering. Vol.3.

[19] Vipin Kumar, Sonajharia Minz, (2014). “Feature Selection: A literature Review”. Smart Computing Review, Vol 4.

[20] Jasmina Novaković, Perica Strbac, Dusan Bulatović, (2011).”Toward Optimal Feature Selection using Ranking Methods and Classification Algorithms”. Yugoslav Journal of Operations Research.

[21] Zena M. Hira, Duncan F. Gillies, (2015) “A Review of Feature Selection and Feature Extraction Methods Applied on Microarray Data”. Hindawi Publishing Corporation Advances in Bioinformatics.

International Journal of Computer Applications Technology and Research
Volume 5– Issue 6, 403 - 411, 2016, ISSN:- 2319–8656

Comparative Study of Optic Fibre and Wireless Technologies in Internet Connectivity

Onu Fergus U.
Department of Computer Science,
Ebonyi State University, Abakaliki – Nigeria

Ikporo Stephen C.
Department of Computer Science,
Ebonyi State University, Abakaliki – Nigeria

Abstract: Most of the activities going on in the world today demand information and data sharing in one form or the other. Consequently, the Internet and its connectivity has gradually become a household concern. The connection to the Internet requires physical transfer of signal (data/information) from one point to another. This can either be through physical medium (wire) or through the air (wireless). This paper a comparative study of Fiber Optics and Wireless Technologies in Internet connectivity seeks to identify which of the two technologies is better for signal transmission in terms of bandwidth utilization, performance, reliability, cost effectiveness, resilience, and security. The study adopted the use of secondary sources for the sourcing of materials. A lot of journal articles, research publications, testbooks, white papers and many more were critically studies and comparatively analysed. It was clear that both media have hitches and challenges. The study showed that although initial cost of acquisition is an inhibitive factor for fibre optic connection, unlimited bandwidth delivery and high Quality of Service (QoS) placed Fiber optics above wireless connectivity in their overall performance.

KEYWORDS: Internet connectivity, quality of service, wireless technology, Bandwidth utilization, Channel Resilience.

1.0 INTRODUCTION

Communication is as old as man himself. Right from creation, man needed to communicate with one another and with their environment. Communication can be verbally or through signs and symbols. This communication no matter the form (verbal or through signs and symbols) is usually through a medium. The medium alone is not the network. The network is the combination of the medium and the communicating devices (active or passive). In every part of the world, leaders and decision makers have recognized the important role which Information and Communication Technology (ICT) has to play in connecting people and as such ICT stimulates and drives employment, economic growth and social development both in the developed and developing countries. For this singular reason, efforts are being made in every part of the world to accelerate the spread of ICT access across all regions.

Despite the improved and substantial investment made in ICT infrastructure in some parts of the world especially Africa in recent years, focus has been on the improvement of the mobile infrastructure and access. Appreciable gaps in the backbone networks are yet to be addressed. This fact has made the effective high-speed Internet services needed for important key business, government and consumer applications continue to be either unavailable or very expensive. In Africa, the cost of broadband Internet access on the average is about three times higher than what is obtainable in Asia, where significant broadband infrastructure investments have been made [1].

The development of broadband access promises an increased economical development. This requires state-of-the-art network connectivity to be realized. Systems like video conferencing, distance education, academic research and remote surgeries, all require large amount of bandwidth, speed, efficiency, great reliability and security. These factors are some of the demands placed on telecommunications networks today [2].

Understanding the best medium to use when trying to transmit information (signal) from one point to another can make such connectivity an easy one. But in the reverse case, this choice can be very troublesome since one needed to understand the performance of the network in terms of the factors such as: bandwidth, performance, reliability, cost efficiency; resiliency, redundancy, and security, which must be considered during Internet connectivity. Generally, [3] categorised network transmission media into guided and unguided groups. Fibre optics belongs to the category of guided media while wireless technology belongs to undguided media family [3] continued.

Most of the recent generations of emerging wireless communication standards utilize improved modulation techniques to squeeze more bandwidth out of available frequency. However, the total bandwidth achieved by wireless technologies, especially the ones using the unlicensed spectrum, are still of magnitude behind what is possible with Fiber optics. Where most unlicensed wireless setups can deliver bandwidths of multiple megabits per second the most advanced Fiber optic connections can deliver multiple gigabits per second, [4].

Ethernet protocol yields gigabit transfer rates and it is increasingly deployed over long distance. With wireless technologies, flexibility in upstream/downstream bandwidth is provided, but high QoS is more difficult to guarantee if bandwidth is shared between users. As a general rule and in application, better QoS is expected from a Fibre optic connection as it provides a dedicated link between two points [4].

This comparative study of optic fiber and wireless technologies in network interconnectivity adopts a methodological technique by considering a set of requirements which include: Application requirements; Technological requirements; Policy and regulatory requirements; Operations and maintenance requirements. The importance of each requirement and each criterion depends on the context of the application. The decision to adopt an optic fiber-based or wireless-based technology for network interconnectivity depends heavily on the evaluation of

these requirements for each technology in the context of the application domain, [5].

With the level of broadband penetration in recent years in all parts of the world, interest has shifted from apenetration to Internet accessibility and efficiency to users. Government and corporate organizations at various levels have made tremendous effort in bringing Internet connectivity to people. In each case, the best medium to achieve this goal of ensuring that people have access to Internet has always been a problem. While some always opt for Fiber optics cable, many others always choose to use wireless connection. This has lead to people trying to fuse the two by using Fiber over Wireless (FoW), and others have tried using Fiber-Wireless (Fi-Wi), [22].

2.0 LITERATURE REVIEW

2.1 Overview of Communication System

Communication system is the system responsible for the transmission of information from one point to another. Communication system can be analogue or digital. Figure 1 shows the block diagram of a typical communication system. The source as represented in figure 1 can generate either analogue signal (eg. speech, audio, image and video), digital data such as text or multimedia data. Information from the source is passed to the source encoder which generates binary data. The generated binary data is modulated to generate waveforms for transmission over the channel. The channel is subjected to various types of impairments (eg. noise) due to its open nature. At the receiver, the whole process is reversed so as to finally restore the original source information, [26].

The communication system can be wired or wireless. Wireless communication deals with the transmission of signal through low-energy radio frequency waves using open air, transmitter and receiver as the media. While wired communication system involves the transmission of signal through wire with the aid of transmitter and receiver. Wired communication system can be through Twisted Pair cable (shielded or unshielded), Coaxial cable or Fiber Optics cable, [24].

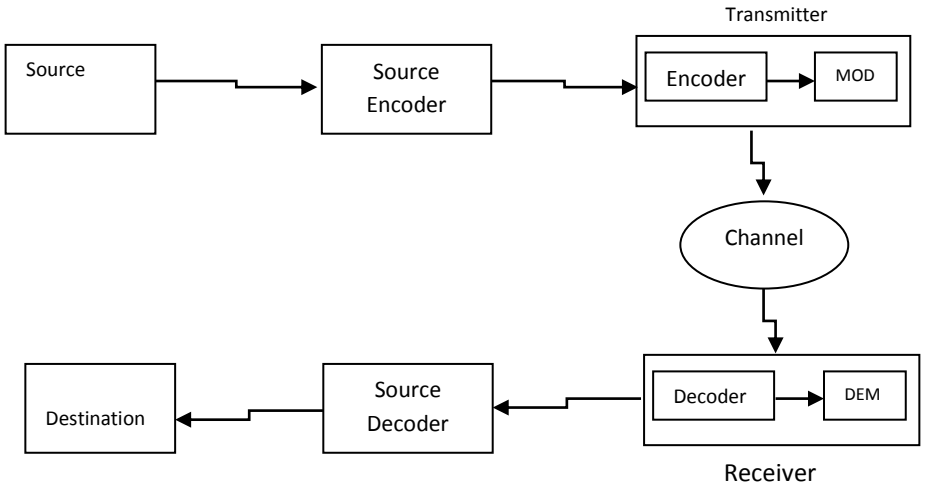


Figure 1: Block Diagram of general communication System

2.2 Wireless Communication System Trends and advancement

This involves the transmission of message signal from one point to another using an open air, transmitter and receiver as the media. This message signal is transmitted to the closest antenna site and delivered by radio signal to another wireless node or receiver, [24]. Wireless communication can be implemented through Wireless Local Area Network (WLAN), Wireless Fidelity (Wi-Fi) etc. The introduction of wireless communication was dated back to 1800s by M.G. Marconi, when he successfully established a radio link between a land-based station and tugboat, [25]. This was succeeded by the invention of Amplitude Modulation (AM) for music broadcasting in 1906 by Fessenden, and in 1933, Edwin H. Armstrong invented Frequency Modulation (FM), [26].

Wireless system advanced rapidly in the last two decades. Wireless communication systems migrated from first-generation (1G) which used narrow-band analogue signalling for voice transmission in 1980s to the second generation (2G) narrow-band systems of 1990s, which used digital communication technique with TDMA, FDMA or CDMA. 2G was deployed for the transmission of voice signal operating on Global System for Mobile Communication (GSM) 900MHz with GPRS 56kbps to 114kbps. 2G also witnessed invention of Global System for Mobility (GSM), Personal Digital cellular (PDC), etc, [4].

Currently, different wireless technologies (eg.GSM, CDMA and TCDMA) were deployed throughout the world for 2G, 2.5G and eventually 3g networks. Despite 2.5G networks offering a higher data rate f 144kbps which was far higher than 2G. It was also used to deliver basic data services like text message. However, 2.5G does not allow for the download of images or even browse a website from PDA, [4].

According to [6], as the limitation of 2G increases, the conception of 3G became more pronounced. The design and deployment of 3G network was aimed at overcoming all the deficiencies of the 2G and 2.5G technologies as it offers more advanced and innovative services such as high-bit rate and broadband for multimedia service.

3G system should be able to prove roaming capabilities and as well be able to make information services instantly available, [7].

3G technology uses digital communication technique that involved transmission of voice signal as well as multimedia services. The Universal Mobile Telecommunication System (UMTS), which is a 3G cell phone technology, was developed into 4G technologies.

4G technology (which is another level in wireless communication) is developed to provide a comprehensive IP solution, where voice, data, and streamed multimedia could be given to user on "anytime" "anywhere" basis, and at a higher data rate than the previous generations. 4G technology is considered to complete the cycle of technological advancements in wireless communications, with features that will permit a much faster velocity in data transfer and wider area coverage than with the current 3G technology, [4].

The earlier wireless systems composed of a base station with a high-power transmitter and covered a significantly large geographic area. Each of the base station served only small number of users and was also very costly. The major technical problem with it was the lack of compatibility in the system as only few were able to communicate with the public switched telephone network (PSTN).

However, today, there have been significant increases in the number of base stations with low-power radio transmitters, which now cover a large geographic area. This is the reason for the proliferation of mobile systems around the globe.

Presently, wireless services have witnessed four stages of progression ranging from simple communication, high-speed downloading, high-speed downloading and uploading, to real-time latency-sensitive services, which are all enabled by the wireless technology. Apart from the multimedia services such as speech, audio, video, and data, the pervasive use of wireless communication could also be used in healthcare, home automation, etc, [4].

2.3 Security Issues in Wireless Communications Systems

Security concern in wireless communication refers to various threats that expose transmitted signals to unauthorised attacks thereby reducing their integrity and confidentiality. In combating this, more sophisticated measures have been taken to ensure that confidentiality and integrity of transmitted information via the wireless media are not compromised. This includes the measures for prevention of unauthorised access or damage to information transmitted over wireless networks. Security issues with wireless communication systems can be grouped into four major components like: physical security, network security, communication security and administrative security. The physical security deals with protection of all facilities where the communication system components are housed. Network security ensures the protection of the system's hardware, software and associated interfaces. The communication security ensures confidentiality and integrity of information transmitted over the air waves. Finally, the administrative security involves the use of procedural control to ensure the confidentiality, integrity and availability of communication systems, [8].
According to [9], [6], [10], and [11], some of the security issues relating to wireless network are summarised as follows:

- **Eavesdropping on Non-Secured Channels:** This is the act of illegitimate interception and reception of information transmitted over a wireless communication system. It is the attack against the confidentiality of data being transmitted over the network. This is possible because it is easy for the network signal to be received outside the vicinity of the valid users. Hence, attackers can hijack the signal over the air from a certain distance, [9].
- **Electromagnetic Interference (EMI):** This occur due to the signal fading and disruption of wireless signals as a result of too long a distance between the transmitter and the receiver, atmospheric conditions or metallic surfaces, which reflect the radio waves and also due to obstacles in the line-of-sight, [10].
- **Overloading of Bandwidth:** This is caused by piggybacking, which is an access to someone else's wireless connection by another within the sphere (area) of the wireless connection without the consent and knowledge of the main subscriber. It can cause service violations, direct attack on ones computer and illegal activities by malicious users, [6].
- **Wireless Sniffing:** This is the invasion of sensitive information like password, bank account numbers, credit card details, etc, using sniffing tools on an unsecured network with an unencrypted traffic. When a network is unsecured, sensitive communication or transactions on them are always at risk.
- **Denial of Service (DoS):** This is the flooding of the network with either valid or invalid messages by the wireless intruders using powerful transceivers enough to cause interference effects on the Wireless LAN (WLAN). They usually cause the WLAN not to be able to communicate using the radio path. DoS causes network interruption and prevent data transmission. This in turn enables malicious attackers observe the recovery of the network and record the codes as they are being re-transmitted by valid user using cracking devices that will aid such attacker to break the security and gain unauthorized access to the system, [11].
- **Spoofing and Session Hijacking:** This is when an attacker impersonates the identity of a valid user to illegitimately gain access to privileged data and resources in the network. This is always the case as 802.11 networks do not authenticate the source address, which is the Media Access Control (MAC) address of the frame. Spoofing can be eliminated by ensuring proper authentication and access control mechanism of WLAN, [11].
- **Traffic Redirection:** This is the manipulation of the MAC address as well as the IP address of a particular wired station by an attacker in order to change the traffic route o f a particular computer to the of the attacker. Others include: Rogue Access Point and Caffe Latte Attack, etc, [10].

2.4 Security Measures with the Wireless Communication

According to [10], 802.11b provided the following security features in wireless network.

- **Service Set Identifier (SSID):** This is the process of ensuring that all devices that require access to a particular WLAN are configured with the same SSID. It is added on the header of the packet sent over the WLAN and are verified by the Access Point. This ensures that clients cannot communicate with a particular access point unless both have same SSID configurations. However, SSID provides little security because it is more of a network identifier than a security feature, [10].

- **Wired Equivalent Privacy (WEP):** This is the standard encryption mechanism for wireless networking. It is an algorithm that is used to protect wireless communications from eavesdropping and modification. It also prevents unauthorized access to a wireless network. WEP relies on secret key shared between a wireless station and an access point which is used to encrypt data packets before they are transmitted and an integrity check is used to ensure the packets are not modified in transit, [13].
- **Media Access Control (MAC) Address Authentication:** Here, the access point is configured to accept association and connection requests from only those modes whose MAC addresses are already registered with it. This provides for additional security layer, [10].

2.5 Wired Communication System

Wired communication refers to the transmission of signals from one point to another through physical media (wire. Wired media is also called guided media [4]. The three identified guided media are Twisted Pair (shielded or unshielded), Coaxial or Fiber Optic Cables.

Twisted Pair Cables (TP): Twisted pair cable is a multi-core (eight) cables twisted into four pairs. Two of the pairs carry the positive or true voltage and are considered tip (T1 to T4), while the other two pairs carry the inverse of voltage grounded and are called ring (R1 to R4). Twisted pair cables are of two categories namely **shielded twisted pair (STP)** and **unshielded twisted pair (UTP).** UTP is the commonest type of cable, but has a problem of crosstalk. STP has sheath for each pair of cables to prevent the interference caused by the crosstalk. TP cable can only reliably connect network segments of not more than 100meters [4].

Coaxial Cable: This consists of a center copper conductor core, a plastic insulating material, a metallic braided sheath, and an outer plastic sheath. The core conducts the data, while the remaining layers provide insulating and protection against signal interferences which might corrupt the transmitted data. A single Coaxial cable has a diameter range of 1 – 2.5cm. It can be used to transmit over a longer distances and support more stations on a shared line than twisted pair, [8]. Coaxial cable has two classes namely: Thinnet (10Base2) and Thicknet (10Base5). **Thinnet or 10Base2** is more flexible than thicknet, hence it is easier to work with. It uses British Naval Connectors (BNC) to attach thinnet cable using T connectors, barrel connectors and terminators. **Thicknet (10base5)** is rarely used today because it is difficult to install and has slow transmission speed [8]. Coaxial cable no matter the type can reliably connect segments of about 185meters.

Fiber Optic Cable: Figure 2 shows a sample optic fibre cable. It is a flexible thin filament of glass (silicon glass) that can accept electrical signals and covert them into optical (light) signals which are reconverted to electrical signals at its destination. They are non-metallic and hence not susceptible to interferences like electromagnetic interference (EMI), radio frequency (RF) or lightning. They are typically smaller and lighter in weight and are practically impervious to outdoor atmospheric conditions. Fiber optic networks are the backbone of the Internet and our enterprise communications infrastructure and can transmit data, video and other applications. They can transmit up to 62 miles before the signals need to be regenerated [19]. The structure of the cable consists of concentric sections shown in figure 2:

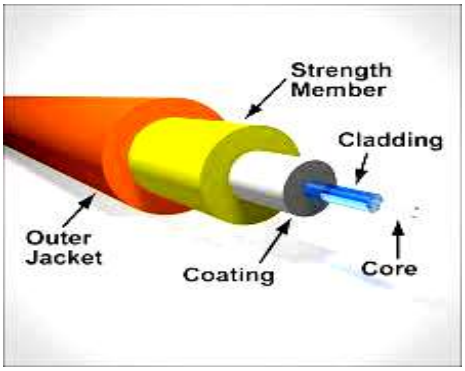


Figure 2: A Sample Optic Fiber cable. [Source: 14].

The Core: The core is the innermost (center) section of the wire and consists of one or more very thin strands, or fibers with a diameter range of 8 to 100µm. It is always made up of glass or plastic. They are the carrier of the optical data signal from the transmitting end to the receiving end.

The Cladding: This is the protective polymer which surrounds the core. The interface between the core and the cladding acts as a reflector to confine light that would otherwise escape the core. It is made up of material that is of lower index of refraction than the core. This is why light is reflected back into the core and the data continues to travel without a loss of light.

The Buffer: This is also carried the coating which helps protect the fiber from physical and environmental damage. It is commonly made of a gel material or a thermoplastic material. The coating is normally stripped away from the cladding to allow termination to an optical transmission system during installation.

The Amor: This layer is usually metallic, rigid, weather proof, and very strong. It serves as physical security measure against outside forces or manipulations.

The Jack: This is the outer layer and always orange in colour. It serves as protection against contaminants, moistures, abrasion, crushing and other environmental dangers, [14].

Fiber optics converts packets of data-images, texts, video, and emails into a stream of light (optical signal). The cable carries the light signal from the transmitter to the receiver, which uses photodiode or photocell to detect the light, and then converts it back to an electrical signal. When the distance increases and becomes longer, an optical regenerator is usually used to boost and regenerate the weakened signal.

Fiber optic cable carries significantly smaller amount of fibers, usually between 2 and 48 fiber strands per bundle and have two cores, as shown in figure 4 below. Generally, one core is used for transmission (TX) and the other core for Reception (RX), [14].

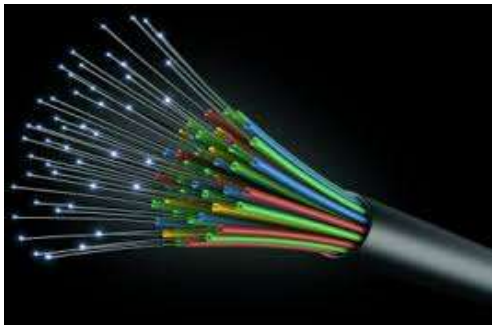


Figure 4: Sample of Optic Fiber with several strands of fiber. [Source: 14].

The signal transmission through Optic fiber is of two modes: Single Mode and Multi Mode;

Single Mode: This is used to transmit signal in longer distances, 50 times more than the multi mode. It has core of between 8-10 micrometer and has a technology that uses powerful laser diode and can transmit with wavelength of 1300/1550nm and a transmission speed range 10GE/1GE/100mbps and a distance of up to 40kmor more. It only has one mode of transmission and costs more than multi-mode, but it is less susceptible to signal attenuation and distortion from overlapping light pulses. The core here is always small, about 9 microns, and also has small numerical aperture (NA).

Multi Mode: This is used to transmit in short and medium distances. It has technology that uses less powerful Light Emitting Diode (LED) which can transmit infrared laser light of wavelengths, 850nm/1300nm. It has core of diameter usually 50, 62.5, or 100 meters with transmission speed of up to 10gbps/1Gbps/100mbps and distance of between 300 meters and 4km. As the name implies, transmission occurs in more than one mode as light waves are dispersed through the cable. Multi-mode fiber can be used with less expensive connectors and LED transmitter, making it more economical choice for application with shorter distances and lower bandwidth demands, [17].

2.6 Security Issues with Fiber Optics

The security issues in fiber optics is basically grouped into: Physical Layer Security and Data Layer Security.

Physical Layer Security: this deals with the physical detection and intrusion to the cable which does not involve significant data interruption. This is why it is not of any advantage to post the fiber optics communication infrastructures on the Internet as it can provide roadmap and bring attention to the fiber optic communications' vulnerabilities. Once an intruder gains access to the cable, the actual tap can easily be done that ever thought or imagined. The intruder can use any available commercial items such as laptop, optical tap, pocket Sniffer Software or even optical/electrical converter to do virtually detectable tap on the cable.

Another physical layer security concern can come from the unintended attackers. This happens when someone unintentionally attacks the cable thereby causing a tap or cut on the cable. Once a successful tap is made, the cable is exposed such that packet sniffer software can be employed to filter through the packet headers.

Data Layer Security: This occurs when the intruder through any tap on the cable tamper with the data being transmitted on the cable. With the cable already tapped, the filter can be applied to the data allowing specified IP addresses, MAC addresses or DNS information to be gathered and then stored or forwarded to the intruding parties' various tools and mechanisms, including other optical connections, links, wireless, another wavelength or other resources, [20]. When the intruder has successfully used an unobtrusive method to retrieve data directly from the fiber optic cable, the need for accessing the company's network will not be necessary. Hence the problem on how to get over firewalls, IDS and IPS will not occur. The only possible problem to the intruder would be when the transmitted data are encrypted. Though depending on the encryption method used, it may still be a matter of time before the intruder breaks the encryption and have their way. Others include;

Optic fiber Splicing: This is the most detectable fiber optical disruptions of data as the cables are cut, thereby allowing for disruption of data transmission. When this type of data interruption/disruption is detected, or noticed, a technician or repair person can be sent out to find the source and fix it.

Optic Fiber Bending or Clamping: This occurs when the cables are tapped without piercing them or disrupting the flow of data. This mainly happen when the cable is bent or clamped in a précised way that can form a micro-bends. When micro-bends or ripples are introduced, photons of light can leak out thereby allowing the intruders' receiver to capture enough of these escaped (leaked) photons of light to have viable data, [20]. This is mostly more pronounced on lowers speed data rates than higher data rates.

According to [20], every signal leak of less than 0.1dB contains all the information being transmitted by each photon. Hence, once the signal is captured, the intruder can use an optical fibre network analyzer to determine the communications protocol and to decipher the information. As far as, there is no disruption or indication of interference with the users' communication the tap is virtually undetected.

This method of tapping fiber cable without actually touching the cable physically, injects additional light into the fiber plant and analyzer, the underlying optical signal protection, an end- user may never notice that their data has been intercepted. Again, another security concern is when intruder gain access to the cable before the first switching center. Detection can go unnoticed even as optical tapping requires less complex and expensive equipment in the local cable and access loops, [20].

2.7 Security Measures in Optic Fiber Communication.

There are some security counter measure in fiber optics communication. These include:

Fibre optic Secure Link: This is used to sense the physical infrastructure disturbance. It uses technology that can concurrently make use of a fibre optic communication cable as a tampering- alert, or integrity- testing sensing cable. It monitors in real-time, any physical disturbance such as clamping or bending. One key advantage this technique is that, it is not necessary for optical losses to occur in order for the technique to sense disturbances, (15).

Encryption Method: This method deals with the encryption of all data being transmitted. This method ensures that a strong encryption with long codes is employed. It also allows for end user to use physical method that can change the light signals as it simultaneously identifies illegal attacks, [16]. Encryption method can also be achieved using photons to encrypt the data. In this method, when a transmitter sends photons that are specifically directed at given intervals through a fiber optic cable, the receiver then analyzes the arrival of the photon at the given intervals. When a matching segment of the transmission pattern advertised on a separate wave length by a transmitter is received, the receiver will then utilize the" key" and authenticate the unlocking of data from the stream. Because of the weakness of the light beam passing through the cable, any alteration would be immediately observed, and any intruder snooping on injecting would inevitably disturb the photon patterns. When this happens, the receiver's device would detect the change in pattern, ending the transmission and sounding the alarm, [17].

Opterna's FiberSentinel System: This uses Wave Sense intrusion prevention technology, artificial intelligence, and optical digital signature recognition to monitor fiber connections. It reportedly detects all physical intrusions and immediately cancels all transmissions. At the time of intrusion detection, this continuous real time monitoring system will switch the data transmission to an alternate fiber path and alerts the network operator, [18].

Oyster Optics Security Solution: This provides optical security, monitoring, and intrusion detection solution that is protocol independent. The system uses a secure phase modulation of the optical signal to impress data on the optical carrier. If data is intercepted, the intruder will not be able to access the captured data unless he/she has Oyster Optics' receiver that is synchronized to the transmitter at power up. This provides a unique transmitter and receiver by using a non-pseudo-random manufacturing process that cannot be replicated. This system can reroute data transmission to a backup system whenever an intrusion is detected. It can be implemented as a stand-alone device or at the transceiver card level, [20].

2.8 How Fiber Optics Communications Work

Fiber Optics technology converts electrical signals carrying data to light and sends the light through transparent glass fibers about the diameter of a human hair. The efficiency of fibre optics is a product of the Index Of Refraction (IOR) and Total Internal Reflection concept. This concept indicated that since fiber is light based, data travels at the speed of light. The speed of light in a vacuum is 186,000 miles per second. When the light is travelling through a medium, the speed is different to it is travelling through a vacuum. The index of refraction is always gotten by dividing the speed of light in a vacuum by the speed of light in a medium. By definition, the IOR of a vacuum has a value of 1. The typical IOR for the core is 1.48 and 1.46 for the cladding. This indicates that light travels slower in the medium, as the IOR gets larger, [21].

The total internal reflection plays key role in the success transmission of data. Total internal reflection occurs when light ray travelling from in one material hits a different material and reflects back in the original material without any loss of light. The core and cladding inside a fiber optic cable work in this manner. The IOR of the core is higher than that of cladding, so when the light from the core hits the cladding, it is reflected back to the core and the data continues to travel. For the total internal reflection to occur, the IOR for the core must be higher than the cladding, [22].

The fibre optics cable's efficiency is also highly enabled by its critical angle. The light of fiber optic must enter through this critical angle. The critical angle of fibre optic is always given by:

$$QC = \cos^{-1}(n2/n1) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (1)$$

Where

n1 is the IOR for the core and
n2 is the IOR for cladding.

E.g. given that n1 is 1.48 and n2 is 1.46, then the critical angle of the given fiber cable,

$$QC = \cos^{-1}(n_2/n_1)$$

$$= \cos^{-1}(1.46/1.48)$$

$$= \cos^{-1}(0.9864864864865)$$

$$= 9.43000^{o}, \text{ [source: 20].}$$

If the angle of incidence is greater than the critical angle, then there will be no angle of refraction. This means that if the light entering the cable hits the core –to – cladding interfaces at an angle greater than the critical angle, it will be reflected back to the core. But if it hits at an angle less than the critical angle, attenuation occurs and the full signal will never reach the receiver.

Attenuation in fibre optics occurs when there is loss to optical power as the light makes its way down the cable. If the light hits impurities in the glass, it will scatter or be absorbed. Extrinsic attenuation may be caused by microbending or macrobending, [21].

Fibre optics use very thin strands of glass or plastic to transmit communication signals. Because they are light based, and data transmitted through them at the speed of light, and they are capable of handling vast amount of data in a much shorter time than copper cable. These light signals use various colours of light (frequencies) as carriers of data. Each colour of light can have multiple hues (sub-frequencies) as separate carriers also, and can carry information for thousands of miles. One strand of fibre carries as much information as 1000 copper cables, making it more efficient and cost effective method of transmitting data over long distances, [19].

2.9 How Wireless Communications Work.

Wireless communication converts data it transmits into electromagnetic waves for broadcasting. Wireless broadband connects homes or business to the internet using a radio link between the customer's location and service provider's station. Wireless technology uses long-range directional equipment to provide broadband service in remote and urban areas with an external antenna being needed. Wireless broadband can be fixed or mobile. The fixed wireless broadband allows the users to access the internet from a fixed point and often require a direct line of sight between the transmitter and the receiver e.g. Wimax802.16d, while the mobile wireless include the Wi-Fi network. Wireless Local Area Network (WLAN) provides wireless broadband access over shorter distances and are often used to extend the reach of wire line or fixed wireless broadband connection within home, organization and campus environment.

Wi-Fi networks use unlicensed devices to provide internet and can be designed for private access for home, or business, or public Internet access as ''Hotspots'' such as campuses, hotels, airports, city parks e.t.c. The Wi-Fi operates at the frequency band of 2.4GHz or 5GHz and support services close to what wired LANs offer (eg. Ethernet). A Wi-Fi enabled device such as a personal computer, mobile phone, etc can connect to the Internet if they are within the range of wireless network connected to Internet. The coverage of one or more interconnected to access point (Hotspot) can be for few rooms or for many square miles covered by access points with overlapping coverage.

Wi-Fi technology can be deployed in mesh configuration, (wireless mesh network) which allows for continuous connection and reconfiguration around broken or blocked paths by ''hopping'' from node to node until the destination is reached. Wi-Fi networks range by design. A typical wireless router using 802.11b or 802.11g with a stock antenna might have a range of 32m (120ft) indoors and 95m (300ft) outdoors. In general, Wi-Fi performance decreases roughly quadratically as distance increases at constant radiation levels. However, by using directional antennas within Line Of Sight (LOS), outdoor ranges can be improved. The Wi-Fi technology can be enhanced and extended by the use of; high gain and MIMO antennas and protocol hacking.

MIMO and High Gain Antennas: A long range Wi-Fi for wireless Metropolitan Area Network uses high gain outdoor directional antennae to establish point-to-point links between fixed points in the system. Using dual antennas with orthogonal polarities along with a 2 x 2 MIMO chipset effectively enable two different carriers' signals to be sent and received along the same long distance path. High gain antenna may be of many designs, but all transmission of narrow signal beam over distance of several kilometres, usually nulling out nearby interference

sources. Another way to extend range is by using power amplifier, commonly known as "range extender amplifiers", which supply around ½ watt of power to the antenna. This amplifier can give more than five times the range of an existing network.

Protocol Hacking: This involves modifying the standard IEEE 802.11 protocol stacks to make them more suitable for long distance, point-to-point usage. This has the risk of breaking interoperability with other Wi-Fi devices and suffes inference from transmitters located near the antenna.

Packet Fragmentation is also used to improve throughput in noisy/congested situations.

3.0 DISCUSSION

This comparative study of fibres optics and wireless technologies in internet connectivity has been presented under two broad topics viz: the similarities and differences between fibre optics and wireless technologies.

Differences between Wireless and Optic Fibre Technologies.

This study presented the difference between Wireless and Optic Fibre Technologies under nineteen (19) key factors as presented in table 1.

Table 1: comparism of wireless and optic fibre technologies

S/	Factor	Optic fibr technology	Wireless technolog
1.	Bandwidth	Bandwidth is bett utilized	Bandwidth is share
2.	Quality of Service (QoS	QoS is better fibre optic	QoS is not as good as in fibre
3.	Efficiency	Fibre is more efficien	Wireless is efficient as fibre
4.	Installation cost	Cost of installing fibre is quiet higher than wireless.	Here wireless has a decisive advantage over fibre
5.	Operation Co	There is recurring costs beyond the installation costs	The extra cost aft installation is less in wireless.
6.	Maintenance Cost	Maintaining connection in fibre cost higher	To maintain connection in wireless is lower
7.	Build–Out Strategies	Adding new access points amounts to extra cost.	New access points can always be added with little or no extra cost.
8.	Network Scalability	Scalability depends on the initial plan of the network	Scalability depends on the load on the base stations and the backhaul
9.	Interference	Fiber uses light waves which is affected by EMI	Wireless uses unlicensed spectrum which is affected by EMI
10	Data Security	Data interception is highly low if not impossible in fiber	Data interception is inherent in wireless technologies.
11	Mobility and Interoperabili	Fiber has interoperability problem because of the wire.	Wireless allows for mobility and interoperability with other devices.
12	Network Coverage	Fiber optics can provide higher network coverage of about 40km or more	Wireless signal degrades with distance and hence can cover a distance of 30 miles.
13	Transmission Speed	Fiber can transmit with a speed as much as 10GE/100m bps	Wireless can transmit with a speed of 30mbps at maximum.
14	Application Support	Fiber can be employed in all layers of a network including access, distribution and backbone	Wireless can only be used as business application and not as a backbone.
15	Health Hazar	Fiber is safer as the light signals transmitted by them are fully contained within the fiber coating with no electromagnetic field or hazardous radiation coming out from them	Wireless technology predisposes people to several diseases like radio frequency sickness, radio wave sickness, microwave sickness, carcinogen, Alzheimer's disease, multiple sclerosis, autism, diabetes, asthma, ADHD, cardiac rhythm and function etc.
16	Environmenta friendliness,	Microwave radiation from wireless can cause reproductive, navigational and health problems for birds, bees, bats and other terrestrial animals. Also too much exposure to Wi-Fi and radiation from antennas causes and barks fissures.	Fiber can cause trending and land breaking, tar cutting, destruction of commercial trees etc especially during installation.
17	Energy Efficiency Mgt.	Fiber optic has better energy efficiency because it uses substantially less energy and does not expose wildlife and vegetation to	Wireless is less energy efficient and can expose wildlife and vegetation to radiation.

		radiation.	
18	Legal and Regulatory Issues	Fiber has legal/regulatory issues pertaining rights of way and pole attachments.	Wireless providers face legal/regulatory issues of unlicensed and licensed spectrum.

4.0 CONCLUSION

This research shows that optic fibre and wireless technologies are complementary to each other. Fiber is always an essential supporting infrastructure for wireless and its real strength lies closer to the core of the network than at the edges. The ultimate goal of any connectivity option should be to have ubiquitous, differentiated network where users can choose the delivery technology that best suits their needs.

Fiber optics provide high data rate in gigabyte per second with good quality link, but it is always expensive when compared to other technologies, and susceptible to many physical (though accidental) attacks. When there is budgeting constraints, wireless technologies, especially WiMax are suitable for implementing a MAN but with a limited data rate of not more than 30 megabyte per second and smaller number of users. The choice of which technology to adopt is definitely and highly depend on the consideration, constraints and priority of the user. For instance, if there is budget or funding constraints the choice of wireless connectivity wiould be favoured, but if there is enough fund, fibre optics is by far a better choice as it provides better quality of service, better bandwidth management and greater number of users.

Both fiber and wireless technologies despite their advantages are still faced with several security issues. Wireless communication network battle with security issues like eavesdropping, wireless sniffing, Denial of Service (DoS), Traffic Redirection, Rogue Access Point, Bandwidth overloading, Electromagnetic Interference, etc, while fibre optics technology have high labour cost, high legal and regulation issue, fiber cut and tapping, etc to contend with.

It is worthy to mention that where users are low and funding is the case including rural areas with difficult terrain; wireless technology is the way to go. However, when there is funding problem and the number of user are high with the network coverage expected to be quite large; fiber optics technology is the answer.

In terms of security of the network, the wireless is inadvertently subjected to various vulnerabilities and security issues. Fiber optics on the other hand always has physical attack and accidental security issues associated with it. Hence, it is almost impossible to have a perfect secured network.

Therefore, the notable practice when making the choice of which technology to implement is always for the need of the user to be placed side by side with the user's priority and constraints.

5.0 REFERENCES

[1.] ARC Electronics. "Brief Overview of Fiber Optic Cable Advantages over copper". The basics of fiber optic cable- an unpublished tutorial. from *http//www.arcelect.com/fibercable.htm* assessed 12/12/2015.

[2.] http://www.businessweek.com/magazine/content/11_09/b4217033849315.htm

[3.] Onu, F. U., Ugwu, I. O. and Okpara C., (2007) Fundamentals of Computer Studies revised edition Copy craft int'l company Ltd. Abakaliki Nigeria pp 224-246.

[4.] http://www.wirtel.co.uk/article_africa_2005_q3_001_alvarion.htm

[5.] Bruno Puzzolante, G. R.c (2006). Nationwide Implementation of a WiMax Mobile Access Network.

[6.] Khan, F., (2009). LTE for 4G Mobile Broadband: Air Interface Technologies and Performance.

[7.] ISO, (2005). "Information Technology-Security Techniques-Code of Practices for Information Security Management.

[8.] Moore, L., (2006). Wireless Technology and Spectrum Demand: Advanced Wireless Services. Congressional Report for Congress.

[9.] Homeland Security, (2006). "Wireless Communication Security"

[10.] Hamid, R. A., (2003).Wireless LAN: "Security Issues and Solutions".

[11.] Sing, G., (2012). Security Issues in Wireless Local Area Network (WLAN).

[12.] US-CERT, (2008). " Using Wireless Technology Securely a government organization". http://www.us-cert.gov/reading_room/home-network-security/> (2014).

[13.] Miller, K. S., (2001). Facing the Challenges of Wireless Security. Technology News **17**

[14.] SANS Institute (2003). An overview of Wireless Security Issues.

[15.] Sovoboda, E., (2005). Code Breakers Stumped by Photon-based System: *Discovery*. **33**

Vol. **26** No. 1.

[16.] Tapanes, E. and Carroll, D., (2010). "Securing Fiber Optics Communication Link against Tapping". Foptic Secure Link- White paper. Available at http/www.fft.com.au/products assessed on 21/12/2015.

[17.] Snawerdt, P., (2002). Phase-Modulated Fiber Optic Telecommunications System.

[18.] Book, E., (2002). "Info-Tech Industry Targets Diverse Threats. Fears of Network Vulnerability fuel market for improving security systems". http://www.americantechsupply.com/fiberopticsecurity.htm.

[19.] Fiber Optic Association, (2004). "Understanding Fiber Optic Communications" available at http:www.thefoa.org/ppt assessed on 12/1/2016.

[20.] Oyster Optics, Inc., (2003). Securing Fiber Optic Communication against Optical Tapping.

[21.] "White Paper on optical taps and various solutions" available at http:www.oysteroptics. com/index resources.html assessed on 3/1/2016.

[22.] Alwayn Virek., (2004). "The Physics behind fiber optics and Fiber Technologies, availabe at

http/www.ciscopress.com/articles/article.asp retrieved on 22/12/2015.

[23.] Corning Incorporated, (2005). "Basic Principles of Fiber Optics", *Corning Cable System:* availabe at http://www.corningcablesystems.com/web/college/fiber tutorial.nsf/apprin?OpenForm retrieved on 11/11/2015.

[24.] Connect Africa Summit 2007, Kigali, Rwanda; availabe at http//www.itu.int/ITU-D/connect/Africa/2007/summit/pdf/s2_background.pdf retrieved on 23/11/2015.

[25.] Finke II, L. G., (2000). "Wireless Communication: A modern Necessity". *Journal of Information Technology* **63**, (*5*).

[26.] Chuah, M., and Zhang, Q., (2006). Design and Performance of 3G Wireless Network and Wireless LANs. USA: Springer.

[27.] Du, K., and Swamy, M. N. S., (2009). Wireless Communication System: From RF Subsystems to 4G Enabling Technologies. London: Cambridge University press.

[28.] Chen, L. and Zheng, L., (2012). "A concentrated-grant-based bandwidth allocation algorithm for Ethernet passive optical networks," in Proceedings of the Symposium on Photonics and Optoelectronics (SOPO '12), pp. 1–4.

International Journal of Computer Applications Technology and Research
Volume 5– Issue 6, 412 - 416, 2016, ISSN:- 2319–8656

Fragmentation of Data in Large-Scale System For Ideal Performance and Security

Dr. Shubhangi D.C
Department of Computer Science and Engineering, VTU Regional Centre Kalaburagi Karnataka, India.

Sonali V.Katke
Department of Computer Science and Engineering, VTU Regional Centre Kalaburagi Karnataka, India.

Abstract: Cloud computing is becoming prominent trend which offers the number of significant advantages. One of the ground laying advantage of the cloud computing is the pay-as-per-use, where according to the use of the services, the customer has to pay. At present, user's storage availability improves the data generation. There is requiring farming out such large amount of data. There is indefinite large number of Cloud Service Providers (CSP). The Cloud Service Providers is increasing trend for many number of organizations and as well as for the customers that decreases the burden of the maintenance and local data storage. In cloud computing transferring data to the third party administrator control will give rise to security concerns. Within the cloud, compromisation of data may occur due to attacks by the unauthorized users and nodes. So, in order to protect the data in cloud the higher security measures are required and also to provide security for the optimization of the data retrieval time. The proposed system will approach the issues of security and performance. Initially in the DROPS methodology, the division of the files into fragments is done and replication of those fragmented data over the cloud node is performed. Single fragment of particular file can be stored on each of the nodes which ensure that no meaningful information is shown to an attacker on a successful attack. The separation of the nodes is done by T-Coloring in order to prohibit an attacker to guess the fragment's location. The complete data security is ensured by DROPS methodology.

Keywords: Performance, Data Splitting, Cloud Security, T-Coloring, Data Security.

1. INTRODUCTION

The usage and management of the information technology infrastructure has reformed by the cloud computing model. Security plays the important aspect to prohibit the prevalent adoption of cloud computing. A cloud contains number of entities in it. In order to provide the security to the cloud, the participating entities have to be secure. Any organization considers the data to be the prime asset. The data must be secured when it is transferred to the public cloud, outside the organizations administrative domain. Accessing the data by the unauthorized users and process should be prevented. Otherwise the weak entity will put the cloud at risk. The prime concern has given to the data availability because of the data may move in the cloud which is not under the customer's administrative control. In cloud computing system, the data reliability problems, data availability problems, response time deal with the strategies of the replication. Data replicas are placed over the number of nodes which increases the attack of the particular data. Here we approach the issue of security and as well as performance as a problem of secure data replication. Within the cloud, the DROPS fragment the files and replicate them at the strategic locations. Based on the user criteria, the files are divided into fragments, and the individual fragment should not contain any meaningful information. In order to increase the data security in the cloud, each of the cloud nodes should contain distinct fragment. An attack on a single node will not show the location of the fragment in the cloud. In order to keep an attacker unsure about the file fragments location and to improve the security, the selection of the nodes is done in such a way that they must not be adjacent and must be at the certain distance from each other. The separation of the nodes is done by T-coloring. The selection of nodes is based on the centrality measures .To improve the retrieval time of the data which will ensure the improved access time. For the selection of the nodes, the two phases are performed. First, based on the centrality measures, the selection of the nodes is done for the initial placing of the fragments. Second, the replication is done for the selected nodes.

2. RELATED WORKS

The problem is the secure and optimal placement of data objects over distributive system within the network. Once the encryption key is divided into n shares and is distributed on different sites. The scheme (k,n) threshold secret sharing scheme is used to divide the key into n shares. The network can be divided into clusters. In every cluster, a primary site is selected which allocates the replicas in it. The scheme used here will combines the problem replication along with the security and improve the access time. This scheme focuses on providing the security to the encryption key. The fragmentation of the data files is not done and handles only single file [17].

The problem, to ensure the freshness, integrity and availability of the data within the cloud is by making use of the cryptographic techniques. The technique depends on the authentication scheme in order to provide the confidentiality of the data. Here the file blocks are stored in the various levels of the tree [6].

Information leakage is another problem that has been seen. In the case of improper sanitization and malevolent VM, the leakage of critical information is not handled. By exploiting the consolidated storage and native access control, the virtualized and multitenancy related issue regarding the cloud storage was approached here [7].

Everyone's perception of software delivery, infrastructure architectures and development models has been drastically altered by while ago emergence of cloud computing. The transition from main frame computer to the client or server implementation models, Cloud Computing includes the elements from grid computing, autonomic and utility computing into innovator implementation architecture. The quick transition towards the cloud has utilized the issue for the result of communication and information security. The Public Key Infrastructure (PKI) is utilized to improve the level of trust in integrity, authentication and confidentiality of data and communication between the involved parties. The certification authorities generate and manage the keys. For the storage of the keys temper proof devices are being utilized such as smart cards [18].

Likewise in [16] the public key cryptography is used and providing the data security in the cloud environment for the trusted third party. But in [16] the PKI infrastructure is not being utilized to minimize the overheads. The generation and management of private and public keys is the responsibility of the trusted third party which may be a single server or multiple servers. Protection can be provided to the symmetric keys by combining the public key cryptography and the (k,n) threshold secret sharing schemes. The dispositioning of the data and loss due to issues arising from virtualization and multi-tenancy cannot be protected under such schemes [16].

The security and optimal placement of data is done through the fragmentation and as well as the data objects replication. The fragmented file is encrypted and is stored within the network in a distributed fashion. In order for the increased availability of the data, replication is performed in a random manner [13].

Architectural and functional blocks of cloud computing which includes the data centers are immanent to the ICT sector (Information Communication Technology). The sundry hosts such as agriculture, nuclear science, and health care, search engines for research, smart grids, data storage and analysis. For cloud under building, a Data Center Networks (DCN) establishes the communicational backbone of a data center, discovering its performance boundaries. In order to overcome the failures, uncertainties in delivering the required Quality of Service level and to satisfy the service level agreement (SLA). The DCN should be strong. In view of manifold failure scenarios for performing a comparative analysis, includes the study of the classical robustness measures and also new procedures to quantitative the robustness of DCN. It also includes presenting multilayer graph modeling of various DCNs and also presents the inadequacy of classical network robustness measures for computing the robustness of the DCN [1].

The system that is an intrusion tolerant distributed system is designed in order that any forcible inclusion into the portion of the system will not endanger the availability, confidentiality and integrity. This approach is fit for distributed system therefore distribution empowers the separation of elements in order that forcible inclusion transfers physical access to only a portion of the system. By forcible inclusion, not only mean that, non-registered people can break-ins computer, but the registered users can also attempts to overstep or invective their immunity. Especially, possible administration security malice is taken into account. Few intensions of distributed systems can be designed to tolerate intrusions. In specific application function such as file management and security function such as user authorization and authentication [2].

The first and foremost aims to significant the major privacy, trust and the security issues in the current existing environment of cloud computing and also help users recognize the intangible and tangible threats associated with their users that includes the surveying for the most applicable security, trust and privacy issues that will pose the threats in the existing environment of cloud computing and also includes for analyzing the way which may be addressed to abolish these trust threats, security and potentiality privacy, and providing a believable, and dependable environment of cloud computing [15].

The proposed scheme that offloads the oftenly occurring dynamic authority generation operation on the trusted entity for keeping the minimum processing responsibility on mobile devices. Moreover, the proposed scheme compares with the existing one on the basis of time, performance metrics and energy consumption [11].

3. SYSTEM ARCHITECTURE

For a large-scale system's security, such as cloud depends on the security of the system and the security of and individual nodes. Thus the issue performance and security as a secure data replication problem is collectively approach. Here, within the cloud the fragment of the data files into pieces is done and replicates those fragments at strategic locations. Based on a user's given criteria, the file is divided into fragments, such that fragments should not include any meaningful information. Every node within the cloud holds a different fragment to raise the security of the data. After all there is the possibility of an effective attack on any node. To keep an attacker unsure about the fragments location and also to improve security, the selection of nodes should be in such a manner that they should not be adjacent and must be at the certain distance from each other. The separation of the nodes is ensured by using T-coloring. Based on the centrality measures, the nodes are selected and the retrieval time of the data is improved that will assure an improved access time. The fragments are replicated over the nodes to improve the retrieval time. The nodes are selected in two stages. In first stage, for initial placing of fragments, the nodes are selected based on the centrality measures. In second stage, for replication the nodes are selected. The data files fragmentation threshold is generated by the file owner. The fragmentation threshold can be specified by the file owner in terms of either percentage or number and size of the different fragments.

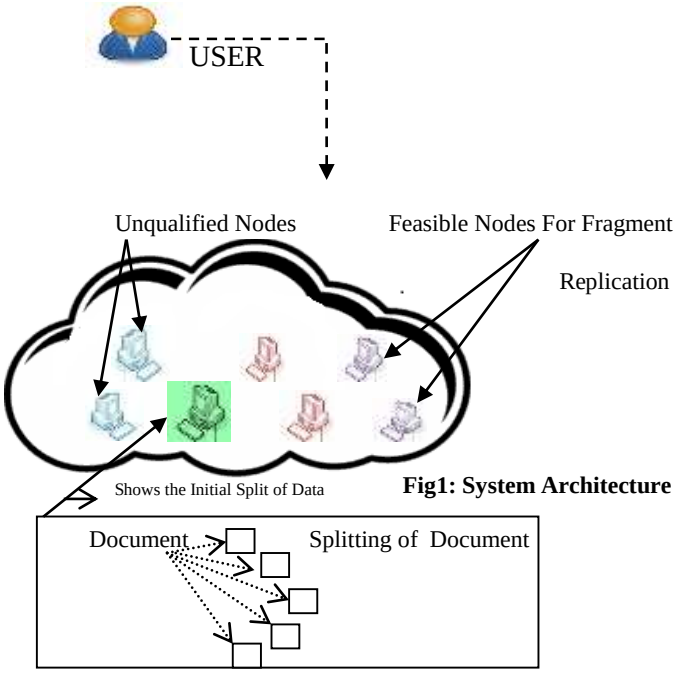


Fig1: System Architecture

For example, the percentage fragmentation threshold can dictate each fragment is of 5% size of the total file size. The owner may create a separate file that containing the information about the size and the fragment number for example, fragment one of size 5000 bytes, and the fragment two of size 8749 bytes. The file is best split by the owner such that each fragment should not contain important information as the owner is aware of the facts that pertaining to the data.

4. METHODOLOGY

To provide the security to the data file within the cloud the entire file is not stored on a single node. The DROPS methodology is used to fragment the file and replicate over the cloud. The fragments should be distributed such that only a node must store a single fragment, though the attack on the node will not be opened for an attacker. In order to improve the security within the cloud, the controlled separation is used by the DROPS methodology in which every fragment is replicated only once. In this methodology, the users need to send the data file to the cloud. Once receiving the data file, the cloud manager system will perform the fragmentation, then selects the node for storing the particular fragment and then replicates the fragments.

Fragment Placement Algorithm

$A = \{A1, A2 \ldots A_N\}$

a = {SIZE OF (A1), SIZE OF (A2)…...SIZE OF (A_N)}

COL = {OPEN_COLOR, CLOSE_COLOR}

CEN = {CEN1, CEN2,…….CEN $_M$}

COL←OPEN_COLOR FOR ALL i

CEN←CEN_i FOR ALL i

COMPUTE:-

FOR EACH A_k BELONGS TO A DO

SELECT P^i/P^i←INDEX OF (MAXIMUM (CEN_i))

IF COL s^i = OPEN_COLOR AND $pi >= a_k$ THEN

$P^i \leftarrow A_k$

$p_i \leftarrow p_i - a_k$

$COL_S{}^i$← CLOSE_COLOR

$P^{i'}$← DISTANCE (P^i,T)

$COL_P{}^i$ ← CLOSE_COLOR

END IF

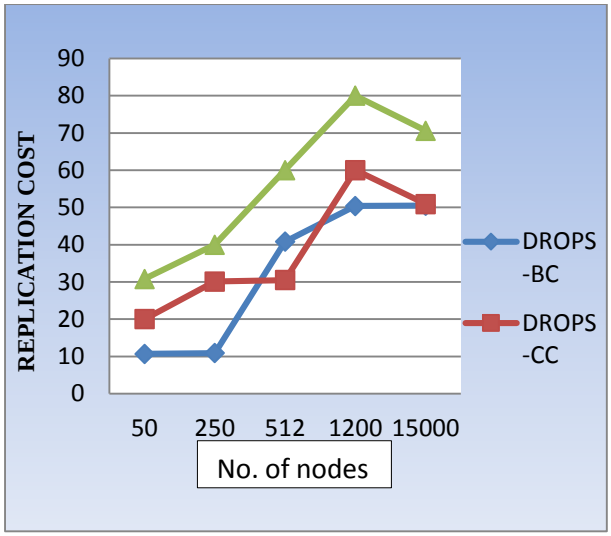
END FOR

5. RESULTS AND DISCUSSIONS

The DROPS methodology performance is compared with the algorithm. The algorithm's behavior is studied by incrementing the number of nodes in the system. By incrementing the number of nodes came to know about the performance of the placement techniques and the DROPS methodology. The Increment in the number of file fragments may strain the cloud storage capacity which in turn can alter the selection of the nodes.

Figure 2: Fragmentation of the data file to provide security.

The data file is divided into fragments, and these fragments are placed on the nodes, in order to provide the security to the data in the cloud.



igure3: Graph represents the centrality measures.

The graph shows the selection of the nodes with the maximum available storage capacity and the maximum centrality. It is obviously true by simple observation that the eccentricity centrality rise in the highest performance although the betweenness centrality resulted the lowest performance. The cause for this is that, in the network the nodes with the higher eccentricity are closer to all different nodes which results in the lower replication cost value for fragments access.

6. CONCLUSION AND FUTUREWORK

For many organizations transferring of data to the wireless servers has become a growing trend, because it takes away the heavy load of the maintenance and the local data storage. This work can consist the problem of generating the copies of data file and store those copies on the cloud servers. In the DROPS methodology, the scheme provided for the security of the cloud storage together as a whole deals with the performance and security in terms of the retrieval time. In order to provide the security to the data file in the cloud, the fragmentation of the data file was performed and the fragments were distributed over many nodes. The separation of nodes is done by using T-coloring. The dispersion and the fragmentation will assure that no important information will be procured by an opponent even on an effective attack. Within the cloud, the node should not store the multiple fragments of the file.

At present, with the methodology of the DROPS the file should be downloaded by the user, update the contents and again upload it. To update and identify the necessary fragments only an automatic update mechanism has to be developed. The time and the resources that are utilized in updating, downloading and again uploading the file will be saved by the above mentioned future work.

ACKNOWLEDGEMENT

We would like to take this opportunity to explicit our deep thankfulness to the Special Officer Dr. Baswaraj Gadge, Head of the Department Dr. Shubhangi D.C and the faculty members of computer science and engineering department of VTU PG CENTRE Kalaburagi for their inspiration and suggestions.

REFERENCES

[1] K. Bilal, M. Manzano, S. U. Khan, E. Calle, K. Li, and A. Zomaya, “On the characterization of the structural robustness of data center networks,” IEEE Transactions on Cloud Computing, Vol. 1, No. 1, 2013, pp. 64-77.
[2] Y. Deswarte, L. Blain, and J-C. Fabre, “Intrusion tolerance in distributed computing systems,” In Proceedings of IEEE Computer Society Symposium on Research in Security and Privacy, Oakland CA, pp. 110-121, 1991.
[3] B. Grobauer, T.Walloschek, and E. Stocker, “Understanding cloud computing vulnerabilities,” IEEE Security and Privacy, Vol. 9, No. 2, 2011, pp. 50-57.
[4] K. Hashizume, D. G. Rosado, E. Fernndez-Medina, and E. B. Fernandez, “An analysis of security issues for cloud computing,”
Journal of Internet Services and Applications, Vol. 4, No. 1, 2013, pp. 1-13.
[5] W. A. Jansen, “Cloud hooks: Security and privacy issues in cloud computing,” In 44th Hawaii IEEE International Conference on System Sciences (HICSS), 2011, pp. 1-10.
[6] A. Juels and A. Opera, “New approaches to security and availability for cloud data,” Communications of the ACM, Vol. 56, No. 2, 2013, pp. 64-73.
[7] G. Kappes, A. Hatzieleftheriou, and S. V. Anastasiadis, “Dike: Virtualization-aware Access Control for Multitenant Filesystems,” University of Ioannina, Greece, Technical Report No. DCS2013-1, 2013.
[8] L. M. Kaufman, “Data security in the world of cloud computing,” IEEE Security and Privacy, Vol. 7, No. 4, 2009, pp. 61-64.
[9] S. U. Khan, and I. Ahmad, “Comparison and analysis of ten static heuristics-based Internet data replication techniques,” Journal of Parallel and Distributed Computing, Vol. 68, No. 2, 2008, pp. 113-136.
[10] A. N. Khan, M. L. M. Kiah, S. U. Khan, and S. A. Madani, “Towards Secure Mobile Cloud Computing: A Survey,” Future Generation Computer Systems, Vol. 29, No. 5, 2013, pp. 1278-1299.
[11] A. N. Khan, M.L. M. Kiah, S. A. Madani, and M. Ali, “Enhanced dynamic credential generation scheme for protection of user identity in mobile-cloud computing, The Journal of Supercomputing, Vol. 66, No. 3, 2013, pp. 1687-1706 .
[12] T. Loukopoulos and I. Ahmad, “Static and adaptive distributed data replication using genetic algorithms,” Journal of Parallel and Distributed Computing, Vol. 64, No. 11, 2004, pp. 1270-1285.
[13] A. Mei, L. V. Mancini, and S. Jajodia, “Secure dynamic fragment and replica allocation in large-scale distributed file systems,” IEEE Transactions on Parallel and Distributed Systems, Vol. 14, No. 9, 2003, pp. 885-896.
[14] L. Qiu, V. N. Padmanabhan, and G. M. Voelker, “On the placement of web server replicas,” In Proceedings of INFOCOM 2001, Twentieth Annual Joint Conference of the IEEE Computer and Communications Societies, Vol. 3, pp. 1587-1596, 2001.
[15] D. Sun, G. Chang, L. Sun, and X. Wang, “Surveying and analyzing security, privacy and trust issues in cloud computing environments,” Procedia Engineering, Vol. 15, 2011, pp. 2852 2856.
[16] Y. Tang, P. P. Lee, J. C. S. Lui, and R. Perlman, “Secure overlay cloud storage with access control and assured deletion,” IEEE Transactions on Dependable and Secure Computing, Vol. 9, No. 6, Nov. 2012, pp. 903-916.

[17] M. Tu, P. Li, Q. Ma, I-L. Yen, and F. B. Bastani, "On the optimal placement of secure data objects over Internet," In Proceedings of 19th IEEE International Parallel and Distributed Processing Symposium, pp. 14-14, 2005.
[18] D. Zissis and D. Lekkas, "Addressing cloud computing security issues," Future Generation Computer Systems, Vol. 28, No. 3, 2012, pp. 583-592.
[19] J. J. Wylie, M. Bakkaloglu, V. Pandurangan, M. W. Bigrigg, S. Oguz, K. Tew, C. Williams, G. R. Ganger, and P. K. Khosla, "Selecting the right data distribution scheme for a survivable storage system," Carnegie Mellon University, Technical Report CMU-CS-01-120, May 2001.
[20] A. R. Khan, M. Othman, S. A. Madani, S. U. Khan, "A survey of mobile cloud computing application models," IEEE Communications Surveys and Tutorials, DOI: 10.1109/SURV.2013.062613.00160.